

PATENT COOPERATION TREATY

PCT

NOTIFICATION OF ELECTION

(PCT Rule 61.2)

From the INTERNATIONAL BUREAU

To:

Assistant Commissioner for Patents  
United States Patent and Trademark  
Office  
Box PCT  
Washington, D.C. 20231  
ÉTATS-UNIS D'AMÉRIQUE

in its capacity as elected Office

Date of mailing: 02 December 1999 (02.12.99)	
International application No.: PCT/JP98/02302	Applicant's or agent's file reference: 98122M
International filing date: 26 May 1998 (26.05.98)	Priority date:
Applicant: ITAI, Akiko et al	

1. The designated Office is hereby notified of its election made:

☒ in the demand filed with the International preliminary Examining Authority on:  
26 May 1998 (26.05.98)

☐ in a notice effecting later election filed with the International Bureau on:

2. The election ☒ was

☐ was not

made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland	Authorized officer:  J. Zahra
Facsimile No.: (41-22) 740 14 35	Telephone No.: (41-22) 338 83 35

PCT

NOTIFICATION OF RECEIPT OF  
RECORD COPY

(PCT Rule 24.2(a))

From the INTERNATIONAL BUREAU

To:

08.06.98

IMAMURA, Masazumi  
Towa Yaesu 1-chome Building  
7th floor  
8-12, Yaesu 1-chome  
Chuo-ku  
Tokyo 103-0028  
JAPON

Date of mailing (day/month/year) 10 June 1998 (10.06.98)	IMPORTANT NOTIFICATION
Applicant's or agent's file reference 98122M	International application No. PCT/JP98/02302

The applicant is hereby notified that the International Bureau has received the record copy of the international application as detailed below.

Name(s) of the applicant(s) and State(s) for which they are applicants:

INSTITUTE OF MEDICINAL MOLECULAR DESIGN, INC. (for all designated States except  
US)

ITAI, Akiko et al (for US)

International filing date : 26 May 1998 (26.05.98)

Priority date(s) claimed :

Date of receipt of the record copy  
by the International Bureau : 09 June 1998 (09.06.98)

List of designated Offices :

AP : GH,GM,KE,LS,MW,SD,SZ,UG,ZW

EA : AM,AZ,BY,KG,KZ,MD,RU,TJ,TM

EP : AT,BE,CH,CY,DE,DK,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE

OA : BF,BJ,CF,CG,CI,CM,GA,GN,ML,MR,NE,SN,TD,TG

National : AL,AM,AT,AU,AZ,BA,BB,BG,BR,BY,CA,CH,CN,CU,CZ,DE,DK,EE,ES,FI,GB,GE,GH,GM,  
GW,HU,ID,IL,IS,KE,KG,KR,KZ,LC,LK,LR,LS,LT,LU,LV,MD,MG,MK,MN,MW,MX,NO,NZ,PL,PT,RO,  
RU,SD,SE,SG,SI,SK,SL,TJ,TM,TR,TT,UA,UG,US,UZ,VN,YU,ZW

ATTENTION

The applicant should carefully check the data appearing in this Notification. In case of any discrepancy between these data and the indications in the international application, the applicant should immediately inform the International Bureau.

In addition, the applicant's attention is drawn to the information contained in the Annex, relating to:

☒ time limits for entry into the national phase;

☒ confirmation of precautionary designations;

☐ requirements regarding priority documents.

A copy of this Notification is being sent to the receiving Office and to the International Searching Authority.

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland	Authorized officer:  M. Sakai
Facsimile No. (41-22) 740.14.35	Telephone No. (41-22) 338.33.33

PCT

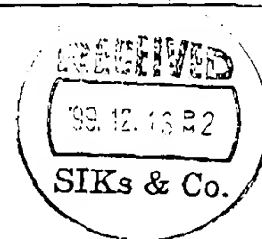
NOTICE INFORMING THE APPLICANT OF THE  
COMMUNICATION OF THE INTERNATIONAL  
APPLICATION TO THE DESIGNATED OFFICES

(PCT Rule 47.1(c), first sentence)

From the INTERNATIONAL BUREAU

To:

IMAMURA, Masazumi  
KRF Building  
5th floor  
5-5, Kyobashi 1-chome  
Chuo-ku  
Tokyo 104-0031  
JAPON



Date of mailing (day/month/year) 02 December 1999 (02.12.99)			
Applicant's or agent's file reference 98122M		IMPORTANT NOTICE	
International application No. PCT/JP98/02302	International filing date (day/month/year) 26 May 1998 (26.05.98)	Priority date (day/month/year)	
Applicant INSTITUTE OF MEDICINAL MOLECULAR DESIGN INC. et al			

1. Notice is hereby given that the International Bureau has communicated, as provided in Article 20, the international application to the following designated Offices on the date indicated above as the date of mailing of this Notice:  
AU,CN,EP,IL,KR,US

In accordance with Rule 47.1(c), third sentence, those Offices will accept the present Notice as conclusive evidence that the communication of the international application has duly taken place on the date of mailing indicated above and no copy of the international application is required to be furnished by the applicant to the designated Office(s).

2. The following designated Offices have waived the requirement for such a communication at this time:  
AL,AM,AP,AT,AZ,BA,BB,BG,BR,BY,CA,CH,CU,CZ,DE,DK,EA,EE,ES,FI,GB,GE,GH,GM,GW,HU,ID,  
IS,KE,KG,KZ,LC,LK,LR,LS,LT,LU,LV,MD,MG,MK,MN,MW,MX,NO,NZ,OA,PL,PT,RO,RU,SD,SE,SG,  
SI,SK,SL,TJ,TM,TR,TT,UA,UG,UZ,VN,YU,ZW  
The communication will be made to those Offices only upon their request. Furthermore, those Offices do not require the applicant to furnish a copy of the international application (Rule 49.1(a-bis)).

3. Enclosed with this Notice is a copy of the international application as published by the International Bureau on  
02 December 1999 (02.12.99) under No. WO 99/62004

**REMINDER REGARDING CHAPTER II (Article 31(2)(a) and Rule 54.2)**

If the applicant wishes to postpone entry into the national phase until 30 months (or later in some Offices) from the priority date, a demand for international preliminary examination must be filed with the competent international Preliminary Examining Authority before the expiration of 19 months from the priority date.

It is the applicant's sole responsibility to monitor the 19-month time limit.

Note that only an applicant who is a national or resident of a PCT Contracting State which is bound by Chapter II has the right to file a demand for international preliminary examination.

**REMINDER REGARDING ENTRY INTO THE NATIONAL PHASE (Article 22 or 39(1))**

If the applicant wishes to proceed with the international application in the national phase, he must, within 20 months or 30 months, or later in some Offices, perform the acts referred to therein before each designated or elected Office.

For further important information on the time limits and acts to be performed for entering the national phase, see the Annex to Form PCT/IB/301 (Notification of Receipt of Record Copy) and Volume II of the PCT Applicant's Guide.

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland	Authorized officer  J. Zahra
Facsimile No. (41-22) 740.14.35	Telephone No. (41-22) 338.83.38

# PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

## INFORMATION CONCERNING ELECTED OFFICES NOTIFIED OF THEIR ELECTION

(PCT Rule 61.3)

To:

IMAMURA, Masazumi  
KRF Building  
5th floor  
5-5, Kyobashi 1-chome  
Chuo-ku  
Tokyo 104-0031  
JAPON



Date of mailing (day/month/year) 02 December 1999 (02.12.99)		
Applicant's or agent's file reference 98122M		IMPORTANT INFORMATION
International application No. PCT/JP98/02302	International filing date (day/month/year) 26 May 1998 (26.05.98)	Priority date (day/month/year)
Applicant INSTITUTE OF MEDICINAL MOLECULAR DESIGN INC. et al		

1. The applicant is hereby informed that the International Bureau has, according to Article 31(7), notified each of the following Offices of its election:

AP : GH, GM, KE, LS, MW, SD, SZ, UG, ZW

EP : AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE

National : AU, BG, BR, CA, CN, CZ, DE, IL, KR, MN, NO, NZ, PL, RO, RU, SE, SK, US

2. The following Offices have waived the requirement for the notification of their election; the notification will be sent to them by the International Bureau only upon their request:

EA : AM, AZ, BY, KG, KZ, MD, RU, TJ, TM

OA : BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG

National : AL, AM, AT, AZ, BA, BB, BY, CH, CU, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IS, KE, KG, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MW, MX, PT, SD, SG, SI, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW

3. The applicant is reminded that he must enter the "national phase" before the expiration of 30 months from the priority date before each of the Offices listed above. This must be done by paying the national fee(s) and furnishing, if prescribed, a translation of the international application (Article 39(1)(a)), as well as, where applicable, by furnishing a translation of any annexes of the international preliminary examination report (Article 36(3)(b) and Rule 74.1).

Some offices have fixed time limits expiring later than the above-mentioned time limit. For detailed information about the applicable time limits and the acts to be performed upon entry into the national phase before a particular Office, see Volume II of the PCT Applicant's Guide.

The entry into the European regional phase is postponed until 31 months from the priority date for all States designated for the purposes of obtaining a European patent.

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland	Authorized officer:  J. Zahra
Facsimile No. (41-22) 740.14.35	Telephone No. (41-22) 338.83.38

P C T

## 国際予備審査報告

(法第12条、法施行規則第56条)  
[PCT36条及びPCT規則70]

出願人又は代理人 の登録記号 98122M	今後の手続きについては、国際予備審査報告の送付通知(様式PCT/ IPEA/416)を参照すること。	
国際出願番号 PCT/JP98/02302	国際出版日 (日.月.年) 26.05.98	優先日 (日.月.年)
国際特許分類 (IPC) Int.Cl <sup>6</sup> G06F17/30		
出願人 (氏名又は名称) 株式会社医薬分子設計研究所		

- 国際予備審査機関が作成したこの国際予備審査報告を法施行規則第57条 (PCT36条) の規定に従い送付する。
- この国際予備審査報告は、この表紙を含めて全部で 4 ページからなる。  
☐ この国際予備審査報告には、附属書類、つまり補正されて、この報告の基礎とされた及び/又はこの国際予備審査機関に対してした訂正を含む明細書、請求の範囲及び/又は図面も添付されている。  
(PCT規則70.16及びPCT実施細則第607号参照)  
この附属書類は、全部で        ページである。
- この国際予備審査報告は、次の内容を含む。
  - ☒ 国際予備審査報告の基礎
  - ☐ 優先権
  - ☒ 新規性、進歩性又は産業上の利用可能性についての国際予備審査報告の不作成
  - ☐ 発明の単一性の欠如
  - ☒ PCT35条(2)に規定する新規性、進歩性又は産業上の利用可能性についての見解、それを裏付けるための文献及び説明
  - ☐ ある種の引用文献
  - ☐ 国際出版の不備
  - ☐ 国際出版に対する意見

国際予備審査の請求書を受理した日 26.05.98	国際予備審査報告を作成した日 09.10.98	
名称及びあて先 日本国特許庁 (IPEA/JP) 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号	特許庁審査官 (権限のある職員)  高 須 功 印	5L 9069
電話番号 03-3581-1101 内線 3564		

1. 国際予備審査報告の基礎

1. この国際予備審査報告は下記の出願書類に基づいて作成された。(法第8条(PCT14条)の規定に基づく命令に  
 応答するために提出された差し替え用紙は、この報告書において「出願時」とする)

☒ 出願時の国際出願書類

- |                              |   |       |      |                      |
|------------------------------|---|-------|------|----------------------|
| <input type="checkbox"/> 明細書 | 第 | _____ | ページ、 | 出願時のもの               |
| 明細書                          | 第 | _____ | ページ、 | 国際予備審査の請求書と共に提出されたもの |
| 明細書                          | 第 | _____ | ページ、 | _____ 付の書簡と共に提出されたもの |
| 明細書                          | 第 | _____ | ページ、 | _____ 付の書簡と共に提出されたもの |
- 
- |                                |   |       |    |                      |
|--------------------------------|---|-------|----|----------------------|
| <input type="checkbox"/> 請求の範囲 | 第 | _____ | 項、 | 出願時に提出されたもの          |
| 請求の範囲                          | 第 | _____ | 項、 | PCT19条の規定に基づき補正されたもの |
| 請求の範囲                          | 第 | _____ | 項、 | 国際予備審査の請求書と共に提出されたもの |
| 請求の範囲                          | 第 | _____ | 項、 | _____ 付の書簡と共に提出されたもの |
| 請求の範囲                          | 第 | _____ | 項、 | _____ 付の書簡と共に提出されたもの |
- 
- |                             |   |       |        |                      |
|-----------------------------|---|-------|--------|----------------------|
| <input type="checkbox"/> 図面 | 第 | _____ | ページ/図、 | 出願時に提出されたもの          |
| 図面                          | 第 | _____ | ページ/図、 | 国際予備審査の請求書と共に提出されたもの |
| 図面                          | 第 | _____ | ページ/図、 | _____ 付の書簡と共に提出されたもの |
| 図面                          | 第 | _____ | ページ/図、 | _____ 付の書簡と共に提出されたもの |

2. 補正により、下記の書類が削除された。

- |                                |   |       |       |
|--------------------------------|---|-------|-------|
| <input type="checkbox"/> 明細書   | 第 | _____ | ページ   |
| <input type="checkbox"/> 請求の範囲 | 第 | _____ | 項     |
| <input type="checkbox"/> 図面    | 第 | _____ | ページ/図 |

3. ☐ この国際予備審査報告は、補充欄に示したように、補正が出願時における開示の範囲を越えてされたものと認めら  
 れるので、その補正がされなかったものとして作成した。(PCT規則70.2(c))

4. 追加の意見(必要ならば)

## Ⅲ. 新規性、進歩性又は産業上の利用可能性についての国際予備審査報告の不作成

次に関して、当該請求の範囲に記載されている発明の新規性、進歩性又は産業上の利用可能性につき、以下の理由により審査しない。

☐ 国際出願全体

☒ 請求の範囲 1 - 4

理由：

☒ この国際出願又は請求の範囲 1 - 4 は、国際予備審査をすることを要しない次の事項を内容としている（具体的に記載すること）。

請求の範囲 1 - 4 は専らその情報の内容に特徴を有する情報の単なる提示と認められるから、特許協力条約に基づく国際出願等に関する法律施行規則第70条第3項で準用する42条（5）に定められた国際予備審査を要しない対象である。

☐ 明細書、請求の範囲若しくは図面（次に示す部分）又は請求の範囲の記載が、不明確であるため、見解を示すことができない（具体的に記載すること）。

☐ 全部の請求の範囲又は請求の範囲が、明細書による十分な裏付けを欠くため、見解を示すことができない。

☐ 請求の範囲について、国際調査報告が作成されていない。

V. 新規性、進歩性又は産業上の利用可能性についての法第12条(PCT35条(2))に定める見解、それを裏付ける文献及び説明

1. 見解

新規性(N)	請求の範囲	5-8	有
	請求の範囲		無
進歩性(IS)	請求の範囲	5-8	有
	請求の範囲		無
産業上の利用可能性(IA)	請求の範囲	5-8	有
	請求の範囲		無

2. 文献及び説明

請求の範囲 5-8

生物学的機能の発現に関する重要度のスコアを付加した情報を含むデータベースから、相同性の評価値を求めることは国際調査に引用されたいずれの文献にも記載されておらず、当業者にとって自明なものでもない。



08.7.23

発信人 日本国特許庁（国際予備審査機関）

出願人代理人

弁理士

今村 正 和

殿

あて名

〒 103-0028

東京都中央区八重洲1丁目8番12号

藤和八重洲一丁目ビル7階

塩澤・今村特許事務所

PCT見解書

(法第13条)

〔PCT規則66〕

発送日

(日.月.年)

21.07.98

出願人又は代理人  
の書類記号

98122M

応答期間

上記発送日から 2 月以内

国際出願番号

PCT/J P 98/02302

国際出願日

(日.月.年)

26.05.98

優先日

(日.月.年)

国際特許分類 (IPC)

Int.Cl. G 06 F 17/30

出願人 (氏名又は名称)

株式会社医薬分子設計研究所

1. これは、この国際予備審査機関が作成した 1 回目の見解書である。
2. この見解書は、次の内容を含む。
- I ☒ 見解の基礎
  - II ☐ 優先権
  - III ☒ 新規性、進歩性又は産業上の利用可能性についての見解の不作成
  - IV ☐ 発明の単一性の欠如
  - V ☒ 法第13条 (PCT規則66.2(a)(ii)) に規定する新規性、進歩性又は産業上の利用可能性についての見解、それを裏付けるための文獻及び説明
  - VI ☐ ある種の引用文獻
  - VII ☐ 国際出願の不備
  - VIII ☐ 国際出願に対する意見
3. 出願人は、この見解書に答えることが求められる。
- いつ?
- 上記応答期間を参照すること。この応答期間に間に合わないときは、出願人は、法第13条 (PCT規則66.2(d)) に規定するとおり、その期間の経過前に国際予備審査機関に期間延長を請求することができる。ただし、期間延長が認められるのは合理的な理由があり、かつスケジュールに余裕がある場合に限り、ことに注意されたい。
- どのように?
- 法第13条 (PCT規則66.3) の規定に従い、答弁書及び必要な場合には、補正書を提出する。補正書の様式及び言語については、法施行規則第62条 (PCT規則66.8及び66.9) を参照すること。
- なお
- 補正書を提出する追加の機会については、法施行規則第61条の2 (PCT規則66.4) を参照すること。補正書及び/又は答弁書の審査官による考慮については、PCT規則66.4の2を参照すること。審査官との非公式の連絡については、PCT規則66.6を参照すること。
- 応答がないときは、国際予備審査報告は、この見解書に基づき作成される。
4. 国際予備審査報告作成の最終期限は、PCT規則69.2の規定により 26.09.00 である。

名称及びあて先

日本国特許庁 (IPEA/J P)

郵便番号 100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

塩澤 正和

5 L

9069

電話番号 03-3581-1101 内線 3584

様式PCT/IPEA/408 (表紙) (1994年1月)

(添付用紙の注意書きを参照)

## I. 見解の基礎

1. この見解書は下記の出願書類に基づいて作成された。(法第8条(PCT14条)の規定に基づく命令に応答するために提出された差替え用紙は、この見解書において「出願時」とする。)

☒ 出願時の国際出願書類

- |                                |         |        |                      |
|--------------------------------|---------|--------|----------------------|
| <input type="checkbox"/> 明細書   | 第 _____ | ページ、   | 出願時に提出されたもの          |
| <input type="checkbox"/> 明細書   | 第 _____ | ページ、   | 国際予備審査の請求書と共に提出されたもの |
| <input type="checkbox"/> 明細書   | 第 _____ | ページ、   | _____ 付の書類と共に提出されたもの |
| <input type="checkbox"/> 請求の範囲 | 第 _____ | 項、     | 出願時に提出されたもの          |
| <input type="checkbox"/> 請求の範囲 | 第 _____ | 項、     | PCT19条の規定に基づき補正されたもの |
| <input type="checkbox"/> 請求の範囲 | 第 _____ | 項、     | 国際予備審査の請求書と共に提出されたもの |
| <input type="checkbox"/> 請求の範囲 | 第 _____ | 項、     | _____ 付の書類と共に提出されたもの |
| <input type="checkbox"/> 図面    | 第 _____ | ページ/図、 | 出願時に提出されたもの          |
| <input type="checkbox"/> 図面    | 第 _____ | ページ/図、 | 国際予備審査の請求書と共に提出されたもの |
| <input type="checkbox"/> 図面    | 第 _____ | ページ/図、 | _____ 付の書類と共に提出されたもの |

2. 補正により、下記の書類が削除された。

- ☐ 明細書 第 \_\_\_\_\_ ページ
- ☐ 請求の範囲 第 \_\_\_\_\_ 項
- ☐ 図面 図面の第 \_\_\_\_\_ ページ/図

3. ☐ この見解書は、補充欄に示したように、補正が出願時における開示の範囲を超えてされたものと認められるので、その補正がされなかったものとして作成した。(PCT規則70.2(c))

4. 追加の意見(必要ならば)

## Ⅲ. 新規性、進歩性又は産業上の利用可能性についての見解の不作成

次に関して、当該請求の範囲に記載されている発明の新規性、進歩性又は産業上の利用可能性につき、次の理由により審査しない。

☐ 国際出願全体

☒ 請求の範囲 1-4

理由:

☒ この国際出願又は請求の範囲 1-4 は、国際予備審査をすることを要しない次の事項を内容としている（具体的に記載すること）。

請求の範囲 1-4 は専らその情報の内容に特徴を有する情報の単なる提示と認められるから、特許協力条約に基づく国際出願等に関する法律施行規則第70条第3項で準用する42条（5）に定められた国際予備審査を要しない対象である。

☐ 明細書、請求の範囲若しくは図面（次に示す部分）又は請求の範囲の記載が、不明確であるため、見解を示すことができない（具体的に記載すること）。

☐ 全部の請求の範囲又は請求の範囲が、明細書による十分な裏付けを欠くため、見解を示すことができない。

☐ 請求の範囲について、国際調査報告が作成されていない。

V. 新規性、進歩性又は産業上の利用可能性についての法第13条（PCT規則66.2(a)(11)に定める見解、それを要する文献及び説明

1. 見解

新規性 (N)

請求の範囲 5-8 有  
請求の範囲 無

進歩性 (IS)

請求の範囲 5-8 有  
請求の範囲 無

産業上の利用可能性 (IA)

請求の範囲 5-8 有  
請求の範囲 無

2. 文献及び説明

請求の範囲 5-8

生物学的機能の発現に関する重要度のスコアを付加した情報を含むデータベースから、相同性の評価値を求めることは国際調査に引用されたいずれの文献にも記載されておらず、当業者にとって自明なものでもない。

12 T  
Translation

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference 98122M	<b>FOR FURTHER ACTION</b> See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)	
International application No. PCT/JP98/02302	International filing date (day/month/year) 26 May 1998 (26.05.98)	Priority date (day/month/year)
International Patent Classification (IPC) or national classification and IPC G06F 17/30		
Applicant INSTITUTE OF MEDICINAL MOLECULAR DESIGN INC.		

<p>1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.</p> <p>2. This REPORT consists of a total of <u>5</u> sheets, including this cover sheet.</p> <p><input type="checkbox"/> This report is also accompanied by ANNEXES, i.e., sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).</p> <p>These annexes consist of a total of _____ sheets.</p>	
<p>3. This report contains indications relating to the following items:</p> <p>I <input checked="" type="checkbox"/> Basis of the report</p> <p>II <input type="checkbox"/> Priority</p> <p>III <input type="checkbox"/> Non-establishment of opinion with regard to novelty, inventive step and industrial applicability</p> <p>IV <input type="checkbox"/> Lack of unity of invention</p> <p>V <input checked="" type="checkbox"/> Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement</p> <p>VI <input type="checkbox"/> Certain documents cited</p> <p>VII <input type="checkbox"/> Certain defects in the international application</p> <p>VIII <input type="checkbox"/> Certain observations on the international application</p>	

Date of submission of the demand 26 May 1998 (26.05.98)	Date of completion of this report 09 October 1998 (09.10.1998)
Name and mailing address of the IPEA/JP	Authorized officer
Facsimile No.	Telephone No.

# INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.

PCT/JP98/02302

## I. Basis of the report

1. This report has been drawn on the basis of *(Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to the report since they do not contain amendments.)*:

- ☒ the international application as originally filed.
- ☐ the description, pages \_\_\_\_\_, as originally filed,  
 pages \_\_\_\_\_, filed with the demand,  
 pages \_\_\_\_\_, filed with the letter of \_\_\_\_\_,  
 pages \_\_\_\_\_, filed with the letter of \_\_\_\_\_.
- ☐ the claims, Nos. \_\_\_\_\_, as originally filed,  
 Nos. \_\_\_\_\_, as amended under Article 19,  
 Nos. \_\_\_\_\_, filed with the demand,  
 Nos. \_\_\_\_\_, filed with the letter of \_\_\_\_\_,  
 Nos. \_\_\_\_\_, filed with the letter of \_\_\_\_\_.
- ☐ the drawings, sheets/fig \_\_\_\_\_, as originally filed,  
 sheets/fig \_\_\_\_\_, filed with the demand,  
 sheets/fig \_\_\_\_\_, filed with the letter of \_\_\_\_\_,  
 sheets/fig \_\_\_\_\_, filed with the letter of \_\_\_\_\_.

2. The amendments have resulted in the cancellation of:

- ☐ the description, pages \_\_\_\_\_
- ☐ the claims, Nos. \_\_\_\_\_
- ☐ the drawings, sheets/fig \_\_\_\_\_

3. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed, as indicated in the Supplemental Box (Rule 70.2(c)).

4. Additional observations, if necessary:

# INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.

PCT/JP98/02302

## III. Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non obvious), or to be industrially applicable have not been examined in respect of:

☐ the entire international application.

☒ claims Nos. 1-4

because:

☒ the said international application, or the said claims Nos. 1-4  
relate to the following subject matter which does not require an international preliminary examination (*specify*):

See supplemental sheet for continuation of Box III

☐ the description, claims or drawings (*indicate particular elements below*) or said claims Nos. \_\_\_\_\_  
are so unclear that no meaningful opinion could be formed (*specify*):

☐ the claims, or said claims Nos. \_\_\_\_\_ are so inadequately supported  
by the description that no meaningful opinion could be formed.

☐ no international search report has been established for said claims Nos. \_\_\_\_\_

**INTERNATIONAL PRELIMINARY EXAMINATION REPORT**

International application No.

PCT/JP 98/02302

**Supplemental Box**

(To be used when the space in any of the preceding boxes is not sufficient)

Continuation of: III. I.

Claims 1-4 are considered to be presentations of information characterized solely by the content of the information, and therefore do not require international preliminary examination, on the basis of Article 42 (5) as applied in Article 70.3 of the rules for enforcing Japanese regulations relating to international applications.



# INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.  
PCT/JP 98/02302

## V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

### 1. Statement

Novelty (N)	Claims	5-8	YES
	Claims		NO
Inventive step (IS)	Claims	5-8	YES
	Claims		NO
Industrial applicability (IA)	Claims	5-8	YES
	Claims		NO

### 2. Citations and explanations

Claims 5-8

Neither of the documents cited in the international search report discloses a database which includes data attributing scores for importance in relation to the expression of biological functions, for seeking evaluations of homology, and this would not be obvious to a person skilled in the art.

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP98/02302

A. CLASSIFICATION OF SUBJECT MATTER  
Int.Cl.<sup>6</sup> G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
Int.Cl.<sup>6</sup> G06F17/30Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched  
Jitsuyo Shinan Koho 1926-1996 Jitsuyo Shinan Toroku Koho 1996-1998  
Kokai Jitsuyo Shinan Koho 1971-1998 Toroku Jitsuyo Shinan Koho 1994-1998Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
JICST File on Science and Technology (DNA, Tanpakushitsu, Aminosan, Heichi)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Katsuki, Kanahisa, "Development of Techniques for Improving the Efficiency of Analyzing Gene Functions (in Japanese)" Report on the Results of Researches for Developing Common Fundamental Technology for Supporting Cancer Research (Period II: fiscal 1987 to fiscal 1989) September 1991 (09. 91), p.49-57, Particularly refer to p.50	5-8
A	Nishioka, Suyama, "Research on the characterization of Amino Acid Sequences on the Basis of Chemical Structures of Ligands (in Japanese)" Report on the Results of Research conducted in fiscal 1993 for Researches concerning Processing of Ample Knowledge Information March 1994 (03. 94), p.56-60	5-8

☐ Further documents are listed in the continuation of Box C.
 ☐ See patent family annex.

* Special categories of cited documents:	T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"U" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Z" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search  
July 7, 1998 (07. 07. 98)Date of mailing of the international search report  
July 21, 1998 (21. 07. 98)Name and mailing address of the ISA/  
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP98/02302

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.: 1-4  
because they relate to subject matter not required to be searched by this Authority, namely:  
The subject matters of claims 1 to 4 are considered mere representations of information characterized solely by the contents of information and thus relate to subject matters which this International Searching Authority is not required, under the provisions of Section 42(5) of the Regulations
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

- Remark on Protest ☐ The additional search fees were accompanied by the applicant's protest.  
☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP98/02302

Continuation of Box No. I of continuation of first sheet (1)

under the Law Concerning International Applications, etc. Pursuant to the Patent Cooperation Treaty, to search.

# INTERNATIONAL SEARCH REPORT

International Application No.

PCT/US 97/19491

A. CLASSIFICATION OF SUBJECT MATTER  
IPC 6 G06F17/30 //C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 0 646 883 A (HITACHI DEVICE ENG ;HITACHI SOFTWARE ENG (JP)) 5 April 1995 see page 4, line 5 - page 8, line 8; claims	1,9,10, 15
A	GB 2 283 840 A (FUJITSU LTD ;NAT INST OF GENERICS (JP)) 17 May 1995 see page 3, line 7 - page 7, line 27; figures 2,7	1,9,10, 15

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

\* Special categories of cited documents :

\*A\* document defining the general state of the art which is not considered to be of particular relevance

\*E\* earlier document but published on or after the international filing date

\*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

\*O\* document referring to an oral disclosure, use, exhibition or other means

\*P\* document published prior to the international filing date but later than the priority date claimed

\*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

\*Z\* document member of the same patent family

Date of the actual completion of the international search

3 April 1998

Date of making of the international search report

21/04/1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel (+31-70) 340-2040, Tx. 31 651 epo nl  
Fax (+31-70) 340-3016

Authorized officer

Fournier, C

## INTERNATIONAL SEARCH REPORT

International Application No.

PCT/US 97/19491

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>DEPETRILLO P B ET AL: "Derivation of a scale-independent parameter which characterizes genetic sequence comparisons" COMPUTERS AND BIOMEDICAL RESEARCH, DEC. 1993, USA, vol. 26, no. 6, ISSN 0010-4809, pages 517-540, XP002061447 see abstract see page 518, paragraph METHODS - page 523, paragraph RESULTS -----</p>	<p>1,9,10, 15</p>

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

# INTERNATIONAL SEARCH REPORT

information on patent family members

International Application No

PCT/US 97/19491

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0646883 A	05-04-95	JP 7093370 A	07-04-95
		US 5706498 A	06-01-98
GB 2283840 A	17-05-95	JP 7274965 A	24-10-95
		US 5598350 A	28-01-97

特許協力条約に基づく出願

願 書

出願人は、この国際出願が特許協力条約に従って処理されることを請求する。

受理官庁記入欄

国際出願番号

国際出願日

(受付印)

出願人又は代理人の書類記号  
(希望する場合、最大12字)

98122M

PCT

28.5.98

440印

控

第 I 欄 発明の名称

蛋白質の機能推定法

第 II 欄 出願人

氏名(名称)及びあて名:(姓・名の順に記載;法人は公式の完全な名称を記載;あて名は郵便番号及び国名も記載)

株式会社医薬分子設計研究所

Institute of Medicinal Molecular Design, Inc.

〒113-0033 日本国東京都文京区本郷5丁目24番5号

角川本郷ビル4F

4th Floor, Kadokawa-hongo Bldg., 24-5,

Hongo 5-chome, Bunkyo-ku, Tokyo 113-0033 JAPAN

☐ この欄に記載した者は、  
発明者でもある。

電話番号:

ファクシミリ番号:

加入電話番号:

国籍(国名): 日本国 JAPAN

住所(国名): 日本国 JAPAN

この欄に記載した者は、次の  
指定国についての出願人である:

☐ すべての指定国

☒ 米国を除くすべての指定国

☐ 米国のみ

☐ 追記欄に記載した指定国

第 III 欄 その他の出願人又は発明者

氏名(名称)及びあて名:(姓・名の順に記載;法人は公式の完全な名称を記載;あて名は郵便番号及び国名も記載)

板井 昭子 ITAI Akiko

〒113-0033 日本国東京都文京区本郷5-16-6

5-16-6, Hongo, Bunkyo-ku,

Tokyo 113-0033 JAPAN

この欄に記載した者は  
次に該当する:

☐ 出願人のみである。

☒ 出願人及び発明者である。

☐ 発明者のみである。  
(ここにレ印を付したとき  
は、以下に記入しないこと)

国籍(国名): 日本国 JAPAN

住所(国名): 日本国 JAPAN

この欄に記載した者は、次の  
指定国についての出願人である:

☐ すべての指定国

☐ 米国を除くすべての指定国

☒ 米国のみ

☐ 追記欄に記載した指定国

☒ その他の出願人又は発明者が続業に記載されている。

第 IV 欄 代理人又は共通の代表者、通知のあて名

次に記載された者は、国際機関において出願人のために行動する:

☒ 代理人

☐ 共通の代表者

氏名(名称)及びあて名:(姓・名の順に記載;法人は公式の完全な名称を記載;あて名は郵便番号及び国名も記載)

9621 弁理士 今村 正純 IMAMURA Masazumi

9263 弁理士 塩澤 寿夫 SHIOZAWA Hisao

9584 弁理士 釜田 淳爾 KAMATA Junji

〒103-0028 日本国東京都中央区八重洲1丁目8番12号

藤和八重洲一丁目ビル7階

7th Floor, Towa Yaesu 1-chome Bldg., 8-12,

Yaesu 1-chome, Chuo-ku, Tokyo 103-0028 JAPAN

電話番号:

03-3271-1331

ファクシミリ番号:

03-3271-1410

加入電話番号:

☐ 代理人又は共通の代表者が選任されておらず、上記枠内に特に通知が送付されるあて名を記載している場合は、レ印を付す



## 第Ⅲ欄の続き その他の出願人又は発明者

この続表を使用しないときは、この用紙を願書に含めないこと。

氏名(名称)及びあて名:(姓・名の順に記載;法人は公式の完全な名称を記載;あて名は郵便番号及び国名も記載)

富岡 伸夫 TOMIOKA Nobuo

〒113-0033 日本国東京都文京区本郷5-21-12  
マンションユキ302号Room 302, Mansion Yuki, 5-21-12, Hongo,  
Bunkyo-ku, Tokyo 113-0033 JAPANこの欄に記載した者は、  
次に該当する:☐ 出願人のみである。☒ 出願人及び発明者である。☐ 発明者のみである。  
(ここに印を付したとき  
は、以下に記入しないこと)

国籍(国名): 日本国 JAPAN

住所(国名): 日本国 JAPAN

この欄に記載した者は、次の  
指定国についての出願人である:☐ すべての指定国☐ 米国を除くすべての指定国☒ 米国のみ☐ 追記欄に記載した指定国

氏名(名称)及びあて名:(姓・名の順に記載;法人は公式の完全な名称を記載;あて名は郵便番号及び国名も記載)

板井 玲子 ITAI Reiko

〒113-0033 日本国東京都文京区本郷5-16-6

5-16-6, Hongo, Bunkyo-ku,  
Tokyo 113-0033 JAPANこの欄に記載した者は、  
次に該当する:☐ 出願人のみである。☒ 出願人及び発明者である。☐ 発明者のみである。  
(ここに印を付したとき  
は、以下に記入しないこと)

国籍(国名): 日本国 JAPAN

住所(国名): 日本国 JAPAN

この欄に記載した者は、次の  
指定国についての出願人である:☐ すべての指定国☐ 米国を除くすべての指定国☒ 米国のみ☐ 追記欄に記載した指定国

氏名(名称)及びあて名:(姓・名の順に記載;法人は公式の完全な名称を記載;あて名は郵便番号及び国名も記載)

今村 正純 IMAMURA Masazumi

〒267-0066 日本国千葉県千葉市緑区あすみが丘4丁目39番地  
ガーデンコート杜の街五番館301号Room 301, Garden-Court Morinomachi-5-bankan,  
39, Asumigaoka 4-chome, Midori-ku, Chiba-shi,  
Chiba 267-0066 JAPANこの欄に記載した者は、  
次に該当する:☐ 出願人のみである。☒ 出願人及び発明者である。☐ 発明者のみである。  
(ここに印を付したとき  
は、以下に記入しないこと)

国籍(国名): 日本国 JAPAN

住所(国名): 日本国 JAPAN

この欄に記載した者は、次の  
指定国についての出願人である:☐ すべての指定国☐ 米国を除くすべての指定国☒ 米国のみ☐ 追記欄に記載した指定国

氏名(名称)及びあて名:(姓・名の順に記載;法人は公式の完全な名称を記載;あて名は郵便番号及び国名も記載)

この欄に記載した者は、  
次に該当する:☐ 出願人のみである。☐ 出願人及び発明者である。☐ 発明者のみである。  
(ここに印を付したとき  
は、以下に記入しないこと)

国籍(国名):

住所(国名):

この欄に記載した者は、次の  
指定国についての出願人である:☐ すべての指定国☐ 米国を除くすべての指定国☐ 米国のみ☐ 追記欄に記載した指定国☐ その他の出願人又は発明者が他の続表に記載されている。

## 第Ⅴ欄 国の指定

規則 4.9 (a) の規定に基づき次の指定を行う (該当する□にレ印を付すこと； 少なくとも1つの□にレ印を付すこと)。

## 広域特許

- ☐ **AP** **ARIPO** 特許: **GH** ガーナ Ghana, **GM** ガンビア Gambia, **KE** ケニア Kenya, **LS** レソト Lesotho, **MW** マラウイ Malawi, **SD** スーダン Sudan, **SZ** スワジランド Swaziland, **UG** ウガンダ Uganda, **ZW** ジンバブエ Zimbabwe, 及びハラレプロトコルと特許協力条約の締結国である他の国
- ☐ **EA** **ユーラシア** 特許: **AM** アルメニア Armenia, **AZ** アゼルバイジャン Azerbaijan, **BY** ベラルーシ Belarus, **KG** キルギスタン Kyrgyzstan, **KZ** カザフスタン Kazakhstan, **MD** モルドヴァ Republic of Moldova, **RU** ロシア連邦 Russian Federation, **TJ** タジキスタン Tajikistan, **TM** トルクメニスタン Turkmenistan, 及びユーラシア特許条約と特許協力条約の締結国である他の国
- ☐ **EP** **ヨーロッパ** 特許: **AT** オーストリア Austria, **BE** ベルギー Belgium, **CH** and **LI** スイス及びリヒテンシュタイン Switzerland and Liechtenstein, **DE** ドイツ Germany, **DK** デンマーク Denmark, **ES** スペイン Spain, **FI** フィンランド Finland, **FR** フランス France, **GB** 英国 United Kingdom, **GR** ギリシャ Greece, **IE** アイルランド Ireland, **IT** イタリア Italy, **LU** ルクセンブルグ Luxembourg, **MC** モナコ Monaco, **NL** オランダ Netherlands, **PT** ポルトガル Portugal, **SE** スウェーデン Sweden, 及びヨーロッパ特許条約と特許協力条約の締結国である他の国
- ☐ **OA** **OAPI** 特許: **BF** ブルキナ・ファソ Burkina Faso, **BJ** ベニン Benin, **CF** 中央アフリカ Central African Republic, **CG** コンゴ Congo, **CI** 象牙海岸 Côte d'Ivoire, **CM** カメルーン Cameroon, **GA** ガボン Gabon, **GN** ギニア Guinea, **ML** マリ Mali, **MR** モーリタニア Mauritania, **NE** ニジェール Niger, **SN** セネガル Senegal, **TD** チャード Chad, **TG** トーゴ Togo, 及びアフリカ知的所有権機構と特許協力条約の締結国である他の国 (他の種類の保護又は取扱いを求める場合には点線の上に記載する)

## 国内特許 (他の種類の保護又は取扱いを求める場合には点線の上に記載する)

- |   |   |
|---|---|
| <input type="checkbox"/> <b>AL</b> アルバニア Albania  | <input type="checkbox"/> <b>LU</b> ルクセンブルグ Luxembourg                                       |
| <input type="checkbox"/> <b>AM</b> アルメニア Armenia  | <input type="checkbox"/> <b>LV</b> ラトヴィア Latvia   |
| <input type="checkbox"/> <b>AT</b> オーストリア Austria   | <input type="checkbox"/> <b>MD</b> モルドヴァ Republic of Moldova                                |
| <input type="checkbox"/> <b>AU</b> オーストラリア Australia  | <input type="checkbox"/> <b>MG</b> マダガスカル Madagascar  |
| <input type="checkbox"/> <b>AZ</b> アゼルバイジャン Azerbaijan  | <input type="checkbox"/> <b>MK</b> マケドニア旧ユーゴスラヴィア The former Yugoslav Republic of Macedonia |
| <input type="checkbox"/> <b>BA</b> ボスニア・ヘルツェゴビナ Bosnia and Herzegovina                        |   |
| <input type="checkbox"/> <b>BB</b> バルバドス Barbados   | <input type="checkbox"/> <b>MN</b> モンゴル Mongolia  |
| <input type="checkbox"/> <b>BG</b> ブルガリア Bulgaria   | <input type="checkbox"/> <b>MW</b> マラウイ Malawi  |
| <input type="checkbox"/> <b>BR</b> ブラジル Brazil  | <input type="checkbox"/> <b>MX</b> メキシコ Mexico  |
| <input type="checkbox"/> <b>BY</b> ベラルーシ Belarus  | <input type="checkbox"/> <b>NO</b> ノルウェー Norway   |
| <input type="checkbox"/> <b>CA</b> カナダ Canada   | <input type="checkbox"/> <b>NZ</b> ニュー・ジーランド New Zealand                                    |
| <input type="checkbox"/> <b>CH</b> and <b>LI</b> スイス及びリヒテンシュタイン Switzerland and Liechtenstein | <input type="checkbox"/> <b>PL</b> ポーランド Poland   |
| <input type="checkbox"/> <b>CN</b> 中国 China   | <input type="checkbox"/> <b>PT</b> ポルトガル Portugal   |
| <input type="checkbox"/> <b>CU</b> キューバ Cuba  | <input type="checkbox"/> <b>RO</b> ルーマニア Romania  |
| <input type="checkbox"/> <b>CZ</b> チェッコ Czech Republic  | <input type="checkbox"/> <b>RU</b> ロシア連邦 Russian Federation                                 |
| <input type="checkbox"/> <b>DE</b> ドイツ Germany  | <input type="checkbox"/> <b>SD</b> スーダン Sudan   |
| <input type="checkbox"/> <b>DK</b> デンマーク Denmark  | <input type="checkbox"/> <b>SE</b> スウェーデン Sweden  |
| <input type="checkbox"/> <b>EE</b> エストニア Estonia  | <input type="checkbox"/> <b>SG</b> シンガポール Singapore   |
| <input type="checkbox"/> <b>ES</b> スペイン Spain   | <input type="checkbox"/> <b>SI</b> スロヴェニア Slovenia  |
| <input type="checkbox"/> <b>FI</b> フィンランド Finland   | <input type="checkbox"/> <b>SK</b> スロヴァキア Slovakia  |
| <input type="checkbox"/> <b>GB</b> 英国 United Kingdom  | <input type="checkbox"/> <b>SL</b> シエラレオネ Sierra Leone                                      |
| <input type="checkbox"/> <b>GE</b> グルジア Georgia   | <input type="checkbox"/> <b>TJ</b> タジキスタン Tajikistan  |
| <input type="checkbox"/> <b>GH</b> ガーナ Ghana  | <input type="checkbox"/> <b>TM</b> トルクメニスタン Turkmenistan                                    |
| <input type="checkbox"/> <b>GM</b> ガンビア Gambia  | <input type="checkbox"/> <b>TR</b> トルコ Turkey   |
| <input type="checkbox"/> <b>GW</b> ギニアビサウ Guinea-Bissau                                       | <input type="checkbox"/> <b>TT</b> トリニダード・トバゴ Trinidad and Tobago                           |
| <input type="checkbox"/> <b>HU</b> ハンガリー Hungary  | <input type="checkbox"/> <b>UA</b> ウクライナ Ukraine  |
| <input type="checkbox"/> <b>ID</b> インドネシア Indonesia   | <input type="checkbox"/> <b>UG</b> ウガンダ Uganda  |
| <input type="checkbox"/> <b>IL</b> イスラエル Israel   | <input type="checkbox"/> <b>US</b> 米国 United States of America                              |
| <input type="checkbox"/> <b>IS</b> アイスランド Iceland   |   |
| <input type="checkbox"/> <b>JP</b> 日本 Japan   | <input type="checkbox"/> <b>UZ</b> ウズベキスタン Uzbekistan                                       |
| <input type="checkbox"/> <b>KE</b> ケニア Kenya  | <input type="checkbox"/> <b>VN</b> ヴィエトナム Viet Nam  |
| <input type="checkbox"/> <b>KG</b> キルギスタン Kyrgyzstan  | <input type="checkbox"/> <b>YU</b> ユーゴスラヴィア Yugoslavia                                      |
| <input type="checkbox"/> <b>KR</b> 韓国 Republic of Korea                                       | <input type="checkbox"/> <b>ZW</b> ジンバブエ Zimbabwe   |
| <input type="checkbox"/> <b>KZ</b> カザフスタン Kazakhstan  |   |
| <input type="checkbox"/> <b>LC</b> セントルシア Saint Lucia   |   |
| <input type="checkbox"/> <b>LK</b> スリ・ランカ Sri Lanka   |   |
| <input type="checkbox"/> <b>LR</b> リベリア Liberia   |   |
| <input type="checkbox"/> <b>LS</b> レソト Lesotho  |   |
| <input type="checkbox"/> <b>LT</b> リトアニア Lithuania  |   |

以下の□は、この様式の施行後に特許協力条約の締結国となった国を指定 (国内特許のために) するためのものである

- ☐ \_\_\_\_\_
- ☐ \_\_\_\_\_
- ☐ \_\_\_\_\_
- ☐ \_\_\_\_\_
- ☐ \_\_\_\_\_
- ☐ \_\_\_\_\_

出願人は 上記の指定に加えて、規則 4.9 (b) の規定に基づき、特許協力条約の下で認められる全ての国の指定を行う。 \_\_\_\_\_ の国の指定を除く。  
ただし、これらの追加される指定が確認を条件としていること、並びに優先日から15月が経過する前にその確認がなされない指定は、この期間の経過時に、出願人によって取り下げられたものとみなされることを宣言する。(指定の確認は、指定を特定する通知の提出と指定手数料及び確認手数料の納付からなる。この確認は、優先日から15月以内に受理官庁へ提出されなければならない。)

## 第VI欄 優先権主張

他の優先権の主張（先の出願）が追記欄に記載されている。

下記の先の出願に基づき優先権を主張する

国名 (その国において又はその国 について先の出願がされた)	先の出願の出願日 (日、月、年)	先の出願の出願番号	先の出願を受理した官庁名 (広域出願又は国際出 願の場合のみ記入)
(1)			
(2)			
(3)			

先の出願の認証書本が、本件国際出願の受理官庁（日本国特許庁）で発行される場合であって、優先権書類送付請求書を本件国際出願に添付するときは、次の□にレ印を付すこと。

☐ 上記（ ）の番号の先の出願のうち、次の（ ）の番号のものについては、出願書類の認証書本を作成し国際事務局へ送付することを、受理官庁（日本国特許庁の長官）に対して請求している。

## 第VII欄 国際調査機関

国際調査機関（ISA）の選択

ISA/J P

先の調査 上記国際調査機関による別の調査（国際・国際型又はその他）が既に実施又は請求されており、可能な限り当該調査の結果を今回の国際調査の基礎とすることを請求する場合に記入する。先の調査に関連する出願（若しくはその翻訳）又は関連する調査請求を表示することにより、当該先の調査又は請求を特定する。

国名（又は広域官庁）

出願日（日、月、年）

出願番号

## 第VIII欄 照合欄

この国際出願の用紙の枚数は次のとおりである。

- |          |      |
|----------|------|
| 1. 願書    | 4 枚  |
| 2. 明細書   | 16 枚 |
| 3. 請求の範囲 | 1 枚  |
| 4. 要約書   | 1 枚  |
| 5. 図面    | 2 枚  |
| 合計       | 24 枚 |

この国際出願には、以下にチェックした書類が添付されている。

- |   |  |
|---|--|
| 1. <input type="checkbox"/> 別個の記名押印された委任状             | 5. <input checked="" type="checkbox"/> 手数料計算用紙               |
| 2. <input type="checkbox"/> 包括委任状の写し                  | <input checked="" type="checkbox"/> 納付する手数料に相当する特許印紙を貼付した書面  |
| 3. <input type="checkbox"/> 記名押印（署名）の説明書              | <input checked="" type="checkbox"/> 国際事務局の口座への振込みを証明する書面     |
| 4. <input type="checkbox"/> 優先権書類（上記第VI欄の（ ）の番号を記載する） | 6. <input type="checkbox"/> 寄託した微生物に関する書面                    |
|   | 7. <input type="checkbox"/> スクレオチド及び／又はアミノ酸配列リスト（フレキシブルディスク） |
|   | 8. <input type="checkbox"/> その他（例えば、優先権書類送付請求書と具体的に記載する）     |

要約書とともに公表する図として 第 図 を提示する（図面がある場合）

## 第IX欄 提出者の記名押印

各人の氏名（名称）を記載し、その次に押印する。

今村正純



塩澤寿夫



釜田淳爾



## 受理官庁記入欄

1. 国際出願として提出された書類の実際の受理の日	2. 図面 <input type="checkbox"/> 受理された <input type="checkbox"/> 不足図面がある
3. 国際出願として提出された書類を補充する書類又は図面であって その後期間内に提出されたものの実際の受理の日（訂正日）	
4. 特許協力条約第11条(2)に基づく必要な補充の期間内の受理の日	
5. 出願人により特定された 国際調査機関 ISA/J P	
6. <input type="checkbox"/> 調査手数料未払いにつき、国際調査機関に 調査用写しを送付していない	

## 国際事務局記入欄

記録原本の受理の日

P C T

手数料計算用紙

願書附属書

受理官庁記入欄

国際出願番号

出願人又は代理人の書類記号

98122M

受理官庁の日付印

出願人

株式会社医薬分子設計研究所

所定の手数料の計算

1. 及び 2. 特許協力条約に基づく国際出願等に関する法律（国内法）  
第 18 条第 1 項第 1 号の規定による手数料（注 1）  
（送付手数料 [T] 及び調査手数料 [S] の合計）

95,000 円 T+S

3. 国際手数料（注 2）

基本手数料

国際出願に含まれる用紙の枚数 24 枚

最初の 30 枚まで

55,000 円 b1

0 × 1,300 =

0 円 b2

30 枚を超える用紙の枚数 用紙 1 枚の手数料

b1 及び b2 に記入した金額を加算し、合計額を B に記入

55,000 円 B

指定手数料

国際出願に含まれる指定数（注 3） 73

11 × 12,700 =

139,700 円 D

支払うべき指定手数料  
の数（上限は 11）  
（注 4）

1 指定当たり  
の手数料  
（円）

B 及び D に記入した金額を加算し、合計額を I に記入

194,700 円 I

4. 納付すべき手数料の合計

T+S 及び I に記入した金額を加算し、合計額を合計に記入

289,700 円

合 計

（注 1）送付手数料及び調査手数料については、合計金額を特許印紙をもって納付しなければならない。

（注 2）国際手数料については、受理官庁である日本国特許庁の長官が告示する国際事務局の口座への振込みを証明する書面を提出することにより納付しなければならない。

（注 3）願書第 V 欄でレ印を付した□の数。

（注 4）指定数を記入する。ただし、11 指定以上は一律 11 とする。

## 明 細 書

### 蛋白質の機能推定法

#### 技術分野

本発明はアミノ酸配列の情報を基にして、そのアミノ酸配列から構成される蛋白質の機能を推定する方法に関するものである。より詳しく言うと、アミノ酸配列の情報を基にして、そのアミノ酸配列から構成される蛋白質の酵素活性などの生物学的機能をコンピューターで利用可能な特定のデータベースを利用して効率的に推定する方法に関する。

#### 背景技術

蛋白質は生体内で生命活動の維持に必須の物質であり、動物や植物に限らず微生物にもさまざまな蛋白質が存在しており、それぞれ固有の機能や役割を担っている。蛋白質をその機能から大別すると、化学反応を触媒する酵素、シグナル伝達物質の受け皿蛋白である受容体、それ自身がシグナルを伝えるシグナル伝達蛋白、及び特定の物質を結合して輸送する蛋白などに分類でき、それぞれが多様な機能によりさらに細分される。例えば、特定の基質の特定部位を還元する酵素や、蛋白質を加水分解する酵素などのように、酵素はそれぞれ固有の反応を触媒している。

蛋白質は主として 20 種のアミノ酸によって構成されており、50～1000 個程度のアミノ酸がさまざまな順序でペプチド結合により鎖状につながったポリペプチド分子である。蛋白質ごとにアミノ酸のつながる順序（アミノ酸配列または 1 次構造と呼ぶ）が異なっており、その結果、それぞれの蛋白質は異なる生理機能を発現できるようになる。すなわち、長いポリペプチド鎖が折り畳まれて立体構造が形成されると、標的分子（酵素基質分子や受容体基質分子など）の捕捉が可能になり、反応に関わる官能基が適切な位置関係に配置されるなど、目的の生物学

的作用の発現に適した場が提供される。それぞれのアミノ酸配列から固有の立体構造が決まり、その立体構造から生物学的機能が決定されることは容易に推定されるが、それらの関係の必然性は未だうまく説明されていない。

蛋白質の研究は、酵素活性を指標として蛋白質を単離・精製し、分子量や構成アミノ酸数、アミノ酸毎の数を決定した後にアミノ酸配列を決定するという古典的な方法に代わって、末端の 20 残基ほどを決定して対応する遺伝子配列を合成した後、蛋白質をコードする遺伝子を釣り上げて遺伝子側から全部のアミノ酸配列が決定する方法が利用されるようになった。これらの研究ではすでに機能がわかっている蛋白質を対象としてきたが、最近では全く逆の順序で研究が行われることも多い。その理由は、遺伝子配列の解析が極めて容易になったために、蛋白質を単離することなく、遺伝子の側からそれがコードするはずの蛋白質のアミノ酸配列を容易に決定できるようになったことにある。

その結果として、生物学的機能がわからないままアミノ酸配列だけが推定できる蛋白質が急増している。蛋白質の生物学的機能は立体構造に基づいて発現されるので、このような機能未知の蛋白質の立体構造を結晶解析や n m r 解析で解析して、その生物学的作用を推定する試みがなされている。しかしながら、このような構造解析には生化学的な使用量よりはるかに大量で、かつ高純度の試料が必要である。また、立体構造からたたちに生物学的機能が推定されるわけではなく、仮に生物学的機能が推定されたとしてもそれが重要なものであるとは限らないので、研究の投資効率が極端に悪いという問題がある。したがって、蛋白質の立体構造を決定する前に、そのアミノ酸配列を有する蛋白質の生物学的作用を推定する方法の開発が切望されている。このような方法が開発されれば、蛋白質研究や遺伝子研究に多大な貢献があるものと期待される。

立体構造と生物学的機能は密接な関係をもっており、機能が既知の蛋白質の立体構造情報は機能メカニズムを説明するだけでなく、さまざまな目的に役に立つ。蛋白質またはリガンド分子との複合体の 3 次元座標はプロテインデータバンク (Brookhaven National Laboratories, U. S. A.) に収められており、世界中で利

用できるようになっている。現在、収録されている構造数は約 5000 程度であるが、生物種の違いやミュータントを除いた独立の蛋白質で考えると約 400~500 程度である。解析技術の普及と蛋白質の単離精製の技術の進歩などから、結晶解析される蛋白質の数は加速度的に増加しつつあるが、現状では立体構造が解明されないままの蛋白質が圧倒的に多い。

結晶解析や n m r 解析によらずに蛋白質の立体構造やリガンド分子との相互作用の様子を推定する方法としてモデリングを利用することができる。アミノ酸配列においてある程度相同性の高い類似蛋白質の立体構造が既に解析されている場合には、その構造を鋳型として利用してモデリングを行うことにより、アミノ酸残基の対応関係に基づいた立体構造を構築できる。この方法は試料を入手する必要がなく、一般にはコンピュータグラフィックス画面上で対話的に行える点で優れた方法である。例えば、一致していないアミノ酸については側鎖を置換することによって行われる。側鎖のコンフォメーションの問題や挿入または欠損アミノ酸の主鎖の問題を有しているものの、推定構造の信頼性はアミノ酸配列の相同性の高さによって決定され、結晶構造とほぼ同様の扱いが可能である。

アミノ酸の種類ができるだけ一致するように 2 種以上の蛋白質のアミノ酸配列間の対応関係をつける方法 (アラインメント) は、モデリングの目的だけでなく、生物種間やファミリー間での類似性や相違点を調べる目的で頻繁に利用されている。アラインメントの手法では、概念的には一方の配列を他方に対して 1 残基ずつずらしながらアミノ酸の一致のスコアが最もよい対応位置を見つけることになるが、実際には、配列間の対応関係の可能性は無限にあるのできめ細かな配慮と繰り返し操作が必要になり、精度の高い結果を得るためには極めて煩雑な作業が必要になる。例えば、一方の配列に挿入や欠損がある場合も多いので単純に全配列でのアミノ酸の一致度をスコアにすることはできず、部分的に良く一致する部分配列を探すことが必要であり、所定の残基数単位 (ウィンドウ) で一致のスコアを算出する必要もある。場合によっては、完全な一致ではなく性質の似たアミノ酸を相同としてスコアを求める必要もある。

一般に、相同性が低い場合にはアラインメントも一義的には決まらず曖昧さが残るという問題を有しているものの、機能が既知の蛋白質群からアミノ酸配列上の類似性の高い蛋白質を探し出すことができるので、アラインメントは現在のところアミノ酸配列しか手がかりのない蛋白質についてその機能を推定する最も簡便な方法である。このような理由から、アラインメントの煩雑な操作をコンピューターを用いて部分的に自動化する試みがなされている。例えば、アミノ酸配列のマッチング法としてFASTA (Pearson, W.R. and Lipman, D.J., Proc. Natl. Acad. Sci. USA, 85, pp.2444-2448, 1988) と BLAST (Altschul, S.F. et al., J. Mol. Biol., 215, pp.403-410, 1990) が知られている。これらの方法は、短い特定の配列が長い対象配列中に存在するか否かを高速に調べるのに適しているものの、長い配列同士を比較する場合や、相同性が低く断片的にしか一致していないような配列を検索対象とする場合には、類似性の判断や類似部分の抽出が著しく困難であり、相同性の判定精度が低い。従って、これらの方法はアラインメントや蛋白質の機能推定の目的には不十分であり、さらに精度に優れた高速な方法の開発が求められていた。

#### 発明の開示

本発明の課題は、アミノ酸配列の情報を基にして、そのアミノ酸配列から構成される蛋白質の生物学的機能を推定する方法を提供することにある。さらに詳しくは、アミノ酸配列の情報のみが利用可能である場合に、生物学的機能がすでに知られている蛋白質のアミノ酸配列との相同性を特定のデータベースを利用してコンピューターにより効率的に検索し、該アミノ酸配列から構成される蛋白質の生物学的機能を正確かつ高速に検索する方法を提供することにある。

本発明者らは上記の課題を解決すべく鋭意努力した結果、下記の特徴を有するデータベースを用いると、アミノ酸配列から蛋白質の機能を極めて高速かつ高精度に推定できることを見出した。

すなわち本発明は、1又は2以上の生物学的機能が知られている蛋白質のアミ



ノ酸配列の情報を含み、該アミノ酸配列を構成する各アミノ酸残基について該生物学的機能の発現に関する重要度のスコアを付加した情報を含むデータベースを提供するものである。

このデータベースは、例えば、アミノ酸配列の相同性に基づいて生物学的機能が未知な蛋白質の機能を推定するために用いることができ、好ましい態様では、生物学的機能が知られている蛋白質のアミノ酸配列の情報として、蛋白質の3次元構造などの立体構造に関する情報が利用可能な蛋白質のアミノ酸配列を用いて、該アミノ酸配列を構成する各アミノ酸残基について該蛋白質とリガンド分子との結合や生物学的機能の発現に関する重要度のスコアが付加されている。これらのデータベースは、一般的には、フロッピーディスク、CD-ROM、磁気テープ、光ディスクなどの種々の記憶用媒体に格納することが可能である。

別の観点からは、上記データベース中の蛋白質（本明細書において「鋳型蛋白質」と呼ぶ場合がある）と生物学的機能が未知のポリペプチド（本明細書において「対象蛋白質」と呼ぶ場合がある）について、それぞれの構成アミノ酸の一致に対して生物学的機能の発現に関する重要度のスコアを考慮した相同性の評価値を求め、該重要度の高い部位の相同性を表わしたアラインメントを作成する方法が本発明により提供される。

この方法の好ましい態様では、生物学的機能の発現に関して重要度が高い連続した2以上のアミノ酸残基を含むグループ配列を用いて、上記データベース中の蛋白質と対象蛋白質について相同性の高い対応関係を検索する工程を含んでいる。また、他の好ましい態様では、データベース中の1の蛋白質と対象蛋白質について上記アラインメントから相同性の最終評価値を得る工程と、データベースに含まれる全蛋白質についての最終評価値から生物学的機能に関して対象蛋白質に最も類似した蛋白質を推定する工程を含んでいる。これらの方法では、蛋白質全体としての相同性は低くても、生物学的機能に関係した部位の相同性が高い蛋白質を効率よく抽出することができ、対象蛋白質の機能を高速かつ高精度に推定できるという特徴を有している。

## 図面の簡単な説明

第1図は、生物学的機能と立体構造がわかっている4種類の蛋白質のアミノ酸配列を構成する各アミノ酸残基について、それぞれの生物学的機能の発現に関する重要度のスコアを付加したアミノ酸配列の情報を示す図である。図中の記号は、それぞれ、大腸菌由来のジヒドロ葉酸還元酵素 (DHFR-EC)、ウシ由来のトリプシン (TRYP)、ウシ由来のリボヌクレアーゼ A (RNAS)、クジラ由来のミオグロビン (MYGL) を示し、アミノ酸残基は1文字表記で示した。

第2図は、対象蛋白質(DHFR-HM)と、該対象蛋白質に対して最も高い評価値SSを与える鑄型蛋白質として抽出されたDHFR-ECとのアラインメントを示した図である。図中の記号は、それぞれ、ヒト由来のジヒドロ葉酸還元酵素(DHFR-HM)及び大腸菌由来のジヒドロ葉酸還元酵素(DHFR-EC)を示し、1段目はDHFR-HMのアミノ酸番号、2段目はDHFR-HMのアミノ酸配列、3段目はDHFR-ECのアミノ酸配列の部分配列、4段目はDHFR-ECのアミノ酸配列の部分配列のアミノ酸番号を示す。

## 発明を実施するための最良の形態

本発明のデータベースは、1又は2以上の生物学的機能が知られている蛋白質のアミノ酸配列の情報を含み、該アミノ酸配列を構成する各アミノ酸残基について該生物学的機能の発現に関する重要度のスコアを付加した情報を含むことを特徴としている。格納の対象となる蛋白質は、例えば、酵素作用や受容体作用などの1又は2以上の生物学的機能が知られており、全アミノ酸配列が知られているものであればいかなるものでもよいが、好ましくは、その蛋白質の立体構造やリカンド結合部位の立体構造がすでに解明若しくは推定されており、または容易に推定可能なものであることが好ましい。データベースには立体構造が明らかにされた蛋白質に関する情報になるべく多く含まれていることが望ましい。例えば、蛋白質またはリガンド分子との複合体の3次元座標はプロテインデータバンク

(Brookhaven National Laboratories, U.S.A.) に収められているが、約 5000 程度(生物種の違いやミュータントを除いた独立の蛋白質で考えると約 400~500 程度)の蛋白質についての情報が利用可能であり、本発明のデータベースの作成に好適に用いることができる。

本発明のデータベースでは、生物学的機能が知られた蛋白質を構成するアミノ酸配列を基にして、各々のアミノ酸残基について該生物学的機能の発現に関する重要度のスコアを付加することを特徴としている。それぞれの蛋白質について格納する情報としては、例えば、蛋白質名、生物種、臓器・器官、サブタイプ、機能の種類、機能の細分類(例えば、酵素については蛋白分解作用や還元作用などの酵素作用)、酵素の分類番号(EC 番号)、酵素反応や生物学的機能に係るリガンド分子(酵素基質、受容体基質、補酵素、金属イオンなど)、立体構造の由来(例えば、X線結晶解析、n m r 解析、同様な生物学的機能を有する類似蛋白質の情報に基づくモデリングによるものなど)、主たる引用文献名、他のデータベースの参照番号、対象部位リガンド、及び全アミノ酸配列などを挙げることができるが、これらに限定されることはなく、適宜の情報を追加又は削除してもよい。

本発明のデータベースは、上記の情報に加えて、該アミノ酸配列を構成する各アミノ酸残基について該生物学的機能の発現に関する重要度のスコアを付加した情報を含んでいる。重要度のスコアは、例えば、重要度が全くないものについてはゼロ、重要度が極めて高いものについては 10 などの数字やその他の記号を与えることにより行われるが、好ましくは、重要度に関連する2以上要素を勘案しながら算出するのが一般的である。本発明のデータベースには、さらに、生物学的機能の発現に寄与がある(スコアがゼロではない)連続したアミノ酸残基配列(1~n)の有無、スコアづけの方式、スコアの合計、蛋白質間でスコアの合計などを規格化のためのスケール因子などの情報を付加することができるが、付加すべき情報はこれらに限定されることはなく、適宜の情報を追加又は削除してもさしつかえない。なお、一個の蛋白質に複数の生物学的機能や複数のリガンド分

子が知られている場合には、それぞれについての情報を格納しておくことが好ましい。

以下、蛋白質のアミノ酸配列を構成する各アミノ酸残基について該生物学的機能の発現に関する重要度のスコアを与える具体的な手法を説明するが、これらの説明は単に例示のためにのみ示されたものであり、いかなる意味においても限定的に解釈してはならない。また、本発明のデータベースはこれらの手段によって製造されたものに限定されることはない。なお、以下の説明においては、重要度のスコアを数値で表わす例について説明しており、重要度が全くないものについてはゼロ、重要度が高くなるにつれて大きな数値を与えるようにしているが、スコア付けの方法がこのような方法に限定されないことも理解すべきである。

(a) 低分子量のリガンド分子を含む蛋白質複合体の結晶解析がなされている場合：

プロテインデータバンクに収納された蛋白質であって、酵素基質、受容体基質、または阻害剤などの低分子量のリガンド分子との複合体の3次元構造が解析されている場合には、リガンド分子からの各アミノ酸残基の距離を求めて、各アミノ酸残基に距離に応じた重要度のスコアを与えることができる。低分子量のリガンド分子としては、例えば、薬理学的に活性な有機化合物、酵素基質、金属イオンなど、いかなるものであってもよい。例えば、リガンド分子のいずれかの原子（例えば、リガンドとなるCa原子やリガンド分子中の側鎖の1原子など）から10Å以内にあるアミノ酸残基には1、8Å以内なら2、6Å以内なら4、10Åより大きい場合には0などのスコアを与えることが可能である。

(b) 蛋白質単独で結晶解析がなされている場合：

リガンド分子を含まずに結晶解析されている蛋白質についても、さまざまな実験から生物学的機能（酵素活性など）に関係する構造領域が推定される場合には、その近傍のアミノ酸残基に対して上記(a)で行ったような距離に応じた数値を与えることが可能である。生物学的機能と立体構造との対応がいない場合においても、生物学的機能が明確である場合には、その蛋白質の立体構造をコンビ

ュータグラフィックス画面上に描いて回転させながら顕著な内孔を探ることによって、生物学的機能に関係するアミノ酸残基を抽出することができる。

(c) 構造保存領域：

同じ生物学的機能を有する蛋白質について、アミノ酸配列の異なるサブタイプや異生物種の蛋白質など2種以上の立体構造の解析結果を利用できる場合には、それらの構造の重ね合わせを行うことにより、立体構造的に保存された領域を抽出することができ、それらの領域に含まれるアミノ酸残基に高い数値を与えることができる。

(d) モデリング：

蛋白質の立体構造が解析されていない場合、実質的に同一の生物学的作用を有することが知られている類縁蛋白質の立体構造を基にして構築したモデリング構造に基づいて、重要度のスコアをつけることが可能である。例えば、受容体サブタイプやアイソザイム、同じファミリーに属する蛋白質、異生物種の同一機能の蛋白質であってアミノ酸配列の相同性が高い場合にはモデリング構造の信頼性が高いことが知られている。重要度のスコアの与え方は、例えば、上記の各手法と同様に行えばよい。

(e) 生化学実験や遺伝子実験：

例えば生化学的な実験などから生物学的機能の発現に重要であることが推定されるアミノ酸残基や、遺伝子的なアミノ酸の変換の実験（ポイントミューテーションなど）から酵素作用などの生物学的機能の発現に必須であると推定されるアミノ酸残基については、高い重要度を与えることができる。例えば、酵素反応においては、リガンド分子との結合などの観点での評価に加えて、触媒的な役割を果たすアミノ酸残基に大きな数値を与えることが可能である。

(f) 高分子リガンド分子である蛋白質：

一般的に、低分子量のリガンド分子との結合が機能に必須である蛋白質はそのリガンド分子と安定に結合するための内孔を有している。一方、例えば蛋白質などの高分子量のリガンド分子が結合する蛋白質では、顕著な内孔をもたず受容体

蛋白質と分子表面で結合することが多く、受容体蛋白質も顕著な内孔を有しないことがある。例えば、サイトカインのようにそれ自身が高分子リガンドとなる場合には、モノクローナル抗体を用いて推定されるエピトープ領域のアミノ酸残基に大きな数値を与えてもよい。

以上に例示したような手法の1種又は2種以上の組み合わせを用い、さらに必要に応じて適宜の手法を追加することによって、生物学的機能が知られている蛋白質のアミノ酸配列を構成する各アミノ酸残基について、該生物学的機能の発現に関する重要度のスコア付けを行い、重要度のスコアが付加されたアミノ酸配列の情報を作成することができる。なお、アミノ酸残基と生物学的機能の発現との関係については、例えば上記(a)の手法などによりその関係が十分に実証されているものに加えて、その関係がある程度推定可能なものなど、種々の基準をスコア付けに利用できることは言うまでもない。

例えば、結晶構造が既知の蛋白質や、生物学的機能の面から類似の立体構造を有することが推定される蛋白質など、なるべく多くの蛋白質について重要度のスコアが付加されたアミノ酸配列の情報を収集し、コンピューターが利用可能な所定の形式で格納して本発明のデータベースを構築することができる。この際、それぞれの蛋白質の情報の量と質に応じて、それぞれの蛋白質について適宜の異なる基準によりスコアを付加してもよい。もっとも、データベース中にスコアづけの方式と蛋白質間の合計スコアの規格化のためのスケール因子を加えることが必要になる場合もある。なお、上記のような情報の入力は、一定の方式に従ってマニュアルで行うことも可能であるが、一般的には、コンピュータグラフィックス画面上で所定のプログラムを用いて行うのが効率的である。

本発明の方法では、上記データベースを用いて、データベース中に情報が格納された鋳型蛋白質と生物学的機能が未知の対象蛋白質について、アミノ酸残基の重要度のスコアから計算した相同性の評価値が最大になるようにアラインメントを作成し、ついで、データベースに含まれる2以上の鋳型蛋白質、好ましくは全部の鋳型蛋白質について同様なアラインメントを作成した後、鋳型蛋白質間で相同

性の評価値の比較を行い、評価値が最も高い鋳型蛋白質を選出することができる。このようにして選出された鋳型蛋白質は、対象蛋白質との立体構造の類似度が高く、実質的に同一の生物学的機能を有する蛋白質であると推定することが可能である。

上記の方法は、一般的には、本発明の上記データベース中の鋳型蛋白質に関する情報を1つ1つ取り出して、対象蛋白質のアミノ酸配列に対してアラインメント作業をすることにより行われる。対象蛋白質のアミノ酸配列の情報が直接利用できる場合にはその情報を入力して用いればよく、対象蛋白質をコードする遺伝子配列の情報のみが利用可能である場合には、その核酸配列の情報から対象蛋白質のアミノ酸配列の情報を翻訳して用いる必要がある。

本発明の好ましい方法の一例として、鋳型蛋白質のアミノ酸配列において生物学的機能への寄与がある連続した2以上のアミノ酸残基（スコアがゼロでない連続したアミノ酸残基）を含むグループ配列を用いて、鋳型蛋白質と対象蛋白質について相同性の高い対応関係を検索する方法を挙げることができる。もっとも、アラインメント作業はこの方法に限定されることはなく、当業者に利用可能ないかなる方法で行ってもよい。

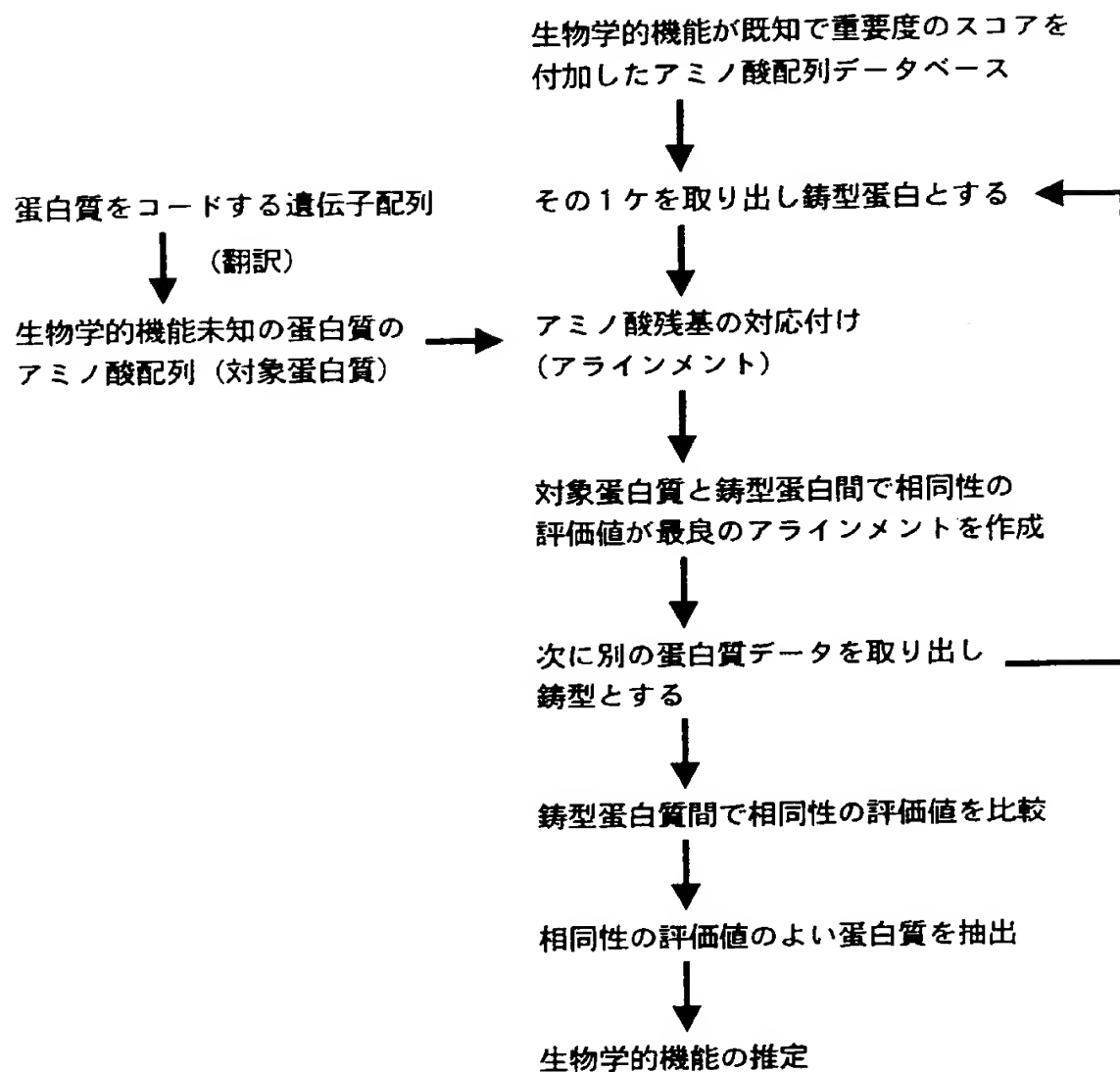
グループ配列を利用する上記の方法では、グループ配列を対象蛋白質のアミノ酸配列に対して1残基ずつずらしながら、グループ毎に相同性の評価値を求めることができ、その後、必要に応じてグループ配列の繋がる順序や長さなどの因子を考慮しつつ、全グループ配列の合計評価値が最良になるように、対象蛋白質のアミノ酸配列に対する各グループ配列の対応関係を決定することができる。この手順をデータベース中のすべての鋳型蛋白質について行い、高い総スコアを有する1又は2個以上の蛋白質を抽出することができる。対象蛋白質は、このようにして抽出された鋳型蛋白質と実質的に同一の生物学的機能を有している可能性が高い。

相同性の評価値としては、例えば、鋳型蛋白質中のグループ配列と対象蛋白質配列中の対応アミノ酸残基が一致した場合には、一致したアミノ酸残基に重要度の

スコアを転記して単に合計するのが簡単である。もっとも、重要度のスコアの高いアミノ酸の一致を重視したアラインメントを作成するためには、重要度の各スコアをさらに1又は2以上の閾値で処理して用いてもよい。対象蛋白質とデータベース中に含まれる全鋳型蛋白質のアラインメントを作成するにあたっては、異なるアラインメント間で全相同性評価値を比較すればよいが、一般的には、大きさの異なる鋳型蛋白質間においても重要度のスコアの合計値によって対象蛋白質との相同性の良し悪しを比較できるように、重要度のスコアの規格化のためのスケール因子を鋳型蛋白質毎に算出してデータベースに格納しておくことが望ましい。対象蛋白質と各鋳型蛋白質とのアラインメント作業が完了して相同性のスコアが計算された段階で、各鋳型蛋白質の相同性スコアに対応のスケール因子を掛け合わせて最終スコアを求め、鋳型蛋白質間での相同性の優劣を決定することができる。

また、異なる生物種に存在する同種蛋白質において、例えば、側鎖の長さが炭素原子1個分違うだけでカルボキシル基を共通に有するアスパラギン酸とグルタミン酸とが、アミノ酸配列中の同じような位置で同一の役割を果たしていることがある。このような場合には、アミノ酸残基の一致を判定するに当たり、これらのアミノ酸残基を一致しているとみなすのが妥当である。また、ロイシンとイソロイシン、バリンなどのアミノ酸残基は、形状やサイズ（嵩高さ）では異なるものの、疎水性の観点からは類似の性質を有している。従って、アミノ酸配列の相同性を数値化する際には、このような類似アミノ酸残基の存在が反映されるように、アミノ酸残基の類似度を段階化した対応表を用いることが望ましい。このようなアミノ酸残基の類似度としてはいかなるものを用いてもよいが、類似度を記載した対応表として、例えば、PAM250 (Dayhoff, M. O., et al., Atlas of Protein Sequence and Structure, Dayhoff, M. O. Ed., Vol. 5, Suppl. 3, pp. 345-352, NBRF, Washington, 1978) や BLOSUM (Henikoff, S. and Henikoff, J. G., Proc. Natl. Acad. Sci. USA, 89, pp. 10915-10919, 1992) などが利用可能である。





本発明の方法の一例を概念図として上記に示した。また、本発明の方法は、例えば下記の工程を含む方法であってもよい。もっとも、本発明の方法はこれらの方法に限定されることはなく、これらの方法において採用された工程に加えて、必要に応じて1又は2以上の適宜の工程を追加することができ、所望であれば1又は2以上の工程を省略できる場合があることを理解すべきである。このような修飾ないし改変された方法がすべて本発明の範囲に包含されることは言うまでもない。

(1) 対象蛋白のアミノ酸配列を呼び出す工程；

- (2) 上記データベースから鋳型配列を1個取り出す工程；
- (3) 鋳型蛋白質のアミノ酸配列から重要度のスコアが一定値以上の部分配列 a, b, c, d, e, ---, n, ---- を例えばN末端から順に取り出す工程（それぞれの部分配列の長さを  $1a, 1b, 1c, 1d, 1e, \dots, 1n, \dots$  とする）；
- (4) 部分配列 a を対象配列の1番目に位置付け、1アミノ酸残基ずつずらしながら相同性の評価値  $S(a)_i$  を算出する工程（相同性の評価値として一致したアミノ酸残基の重要度のスコアを加える）；
- (5) 部分配列 b を対象配列の  $(1+1a)$  番目に位置づけ、1アミノ酸残基ずつずらしながら、相同性の評価値  $S(b)_i$  を算出する工程（相同性の評価値として一致したアミノ酸残基の重要度のスコアを加える）；
- (6) 同様に c, d, e, ---, n, ---- について相同性の評価値  $S(n)_i$  を算出する工程；
- (7) a, b, c, d, e, ---, n, ---- の順序とそれぞれのアミノ酸残基数を考慮しながら、全部分配列の相同性  $SS$  がもっとも大きくなるように部分配列の対応位置を決定する工程；
- (8)  $SS$  にスケール因子をかけて  $SSS$  とする工程；
- (9) データベース中の全鋳型蛋白質について上記の手順を行い  $SSS$  を求める工程；及び
- (10)  $SSS$  の高い蛋白質を抽出する工程

#### 実施例

本発明を下記の実施例により更に具体的に説明するが、本発明の範囲は下記の実施例に限定されることはない。

#### 例1：データベースの作成

生物学的機能と立体構造がわかっている大腸菌由来のジヒドロ葉酸還元酵素 (DHFR-EC)、ウシ由来のトリプシン (TRYP)、ウシ由来のリホヌクレアーゼ A

(RNAS)、クジラ由来のミオグロビン (MYGL) の4種類の蛋白質からなるデータベースを作成した。それぞれの結晶構造はプロテインデータバンク (Brookhaven National Laboratories, U.S.A.) から入手した。アミノ酸残基のいずれかの構成原子が、蛋白質に結合しているリガンド分子 (阻害剤又は補酵素) のいずれかの原子から 4 Å 以内にある場合には2、4~10 Å の範囲内である場合には1、その他の場合には0を与えて、それぞれのアミノ酸配列を構成する各アミノ酸残基についてそれぞれの生物学的機能の発現に関する重要度のスコアを付加したアミノ酸配列の情報を作成した。図1に各蛋白質のアミノ酸配列とスコア付けの結果を示す。

## 例2：生物学的機能の推定

ヒト由来のジヒドロ葉酸還元酵素 (DHFR-HM) を対象蛋白質とし、上記データベースを用いて、該対象蛋白質の生物学的機能を本発明の方法により推定した。DHFR-HM は生物学的機能及び立体構造が知られている蛋白質であるが、生物学的機能及び立体構造を未知なものとして解析を行った。データベース中の鋳型蛋白質のそれぞれのアミノ酸配列について、スコア値が1以上の部分配列を取り出し、対象蛋白質のアミノ酸配列に対して1残基ずつずらしながら相同性の評価値Sを算出し、評価値Sの最も高いアラインメント位置を決定した。

評価値Sの算出には、アミノ酸の類似度に関する対応表 BLOSUM62 を用い、部分配列と対象蛋白質のアミノ酸配列とのアミノ酸残基の組に応じた類似度の評価値を表から求めて、部分配列の各残基のスコアと類似度の評価値との積を部分配列の長さにわたって積算する方法を採用した。全部分配列の相同性SSは、部分配列ごとに評価値Sの最大値を求めて、それらの和として算出した。相同性決定に用いた配列の長さの違いを補正するために、スケール因子として全部分配列のスコアの総和の逆数を用いて、スケール因子をSSの値に乘して最終的な評価値SSSを算出した。この結果、表1に示すように、DHFR-EC, TRYP, RNAS, MYGL を鋳型蛋白質として用いた場合、DHFR-EC が最も高い評価値SSSを与え、対象蛋

白質(DHFR-HM) が DHFR-EC と類似の蛋白質であり、ジヒドロ葉酸還元酵素の活性を有するものと推定された。図2には DHFR-HM と DHFR-EC のアラインメントを示した。

表 2

蛋白質	評価値 SSS
DHFR-EC	1.82
TRYP	1.09
RNAS	1.22
MYGL	0.61

#### 産業上の利用可能性

本発明のデータベースは、アミノ酸配列の情報を基にして、そのアミノ酸配列から構成される蛋白質の生物学的機能を推定するために有用であり、本発明の方法は、このデータベースを利用してアミノ酸配列から構成される蛋白質の生物学的機能を正確かつ高速に検索することができるので有用である。

## 請 求 の 範 囲

1. 1又は2以上の生物学的機能が知られている蛋白質のアミノ酸配列の情報を  
含み、該アミノ酸配列を構成する各アミノ酸残基について該生物学的機能の発現  
に関する重要度のスコアを付加した情報を含むデータベース。
2. アミノ酸配列の相同性に基づいて生物学的機能が未知な蛋白質の機能を推定  
するために用いる請求の範囲第1項に記載のデータベース。
3. 生物学的機能が知られている蛋白質のアミノ酸配列の情報として、蛋白質の  
立体構造に関する情報が利用可能な蛋白質のアミノ酸配列を用いて作成された請  
求の範囲第1項又は第2項に記載のデータベース。
4. 記憶用媒体に格納された請求の範囲第1項ないし第3項のいずれか1項に記  
載のデータベース。
5. 請求の範囲第1項ないし第3項のいずれか1項に記載のデータベースに格納  
された蛋白質及び生物学的機能が未知のポリペプチドについて、それぞれの構成  
アミノ酸の一致に対して生物学的機能の発現に関する重要度のスコアを考慮した  
相同性の評価値を求め、該重要度の高い部位の相同性を表わしたアラインメント  
を作成する方法。
6. 生物学的機能の発現に関して重要度が高い連続した2以上のアミノ酸残基を  
含むグループ配列を用いて、上記データベース中の蛋白質及び対象蛋白質につい  
て相同性の高い対応関係を検索する工程を含む、請求の範囲第5項に記載の方法。
7. データベース中の1の蛋白質と対象蛋白質について上記アラインメントから  
相同性の最終評価値を得る工程を含む、請求の範囲第5項又は第6項に記載の方  
法。
8. データベースに含まれる全蛋白質についての最終評価値から生物学的機能に  
関して対象蛋白質に最も類似した蛋白質を推定する工程を含む、請求の範囲第7  
項に記載の方法。

## 要 約 書

1 又は 2 以上の生物学的機能が知られている蛋白質のアミノ酸配列の情報を含み、該アミノ酸配列を構成する各アミノ酸残基について該生物学的機能の発現に関する重要度のスコアを付加した情報を含むデータベース、及び該データベースに格納された蛋白質及び生物学的機能が未知のポリペプチドについて、それぞれの構成アミノ酸の一致に対して生物学的機能の発現に関する重要度のスコアを考慮した相同性の評価値を求め、該重要度の高い部位の相同性を表わしたアラインメントを作成する方法。

# DHFR-EC

10 20 30 40 50 60 70 80 90 100  
 MISLIAALAVDRVIGMENAMPWNLPADLAWFKRNTLDKPVIMGRHTWESIGRPLPGRKNIILSSQPGTDDRVTWVKSVDIAAAGCDVPEIMVIGGGRVY  
 1122211 112221122211111 11111211221212112 112111

110 120 130 140 150 159  
 EQFLPKAQKLYLTHIDAEVEGDTHFPDVEPDWESVFSEFHDADAQNSHSCFKILERR  
 111211

# TRYP

10 20 30 40 50 60 70 80 90 100  
 I'CGYTCGANTVEYQVELNSGYHEFCGGGLINSQWVVSAAHCTYKSGIQVFLGEDNINIVVEGNSQFISASKSIVHPSTNSHTLNNDIMLIRLKSAAASLNSRV  
 11111 11111

110 120 130 140 150 160 170 180 190 200  
 ASISLPTSCASAGTQCLISGWENTKSSGTSYFDVLKCLKAFILSDSSCKSAVFGQITSNMFCAGYLEGGRKDSQQGDSGPGPVVCSGKLQGGIVSWSGGCAQK  
 111111111 11111 1122221121111 11222222221

210 220  
 NPGYVYTKVGNVVSNIKQTIASN  
 122221

# RNAS

10 20 30 40 50 60 70 80 90 100  
 PETAAAPFERQCHMDSSSTAASSSNYCNQNKKEFNLTEDFCRFPVNTFVHESLADVQAVCSQKNVACNGGQTHCYQSYSTMSITDRETGSSSKYFNCAHYFTT  
 11111111221 2122211 111121

110 120 124  
 QAKKHIIIVACEGNEFVVFHEDASV  
 1111111 111222211

# MYGL

10 20 30 40 50 60 70 80 90 100  
 VLSEGEWQIVLHVWAKVEADVAGHGQDILIRLFKSHFETLEKFFDFKHLKTEAEMKASEDLKKKHGVTVLIALGAILKKKSHHEAELKPLAQSHATKHKIP  
 11111 11121122121 111211221 12121

110 120 130 140 150  
 IKYLEFISEAIHVLHSRHPGDFGADAQGAMNKALELFRKDIKAKYKELGYQG  
 11222121111 1111211111111

第2図

DDFR-HM	10	20	30	40	50	60	70
	VGSLNCIVAVSQNMGIGKNGDL	PWPP	LNFRNFRYFQ	MTTSS	VEGKQNL	VIMGKKTWF	SIPEKNRPLKGR
DDFR-EC	SLIAALA		LPADLAWFKRNTL	DK		VIMGRHTWESI	GRPLPGR
	3	9	24	38	40		57
	80	90	100	110	120	130	140
DDFR-HM	INLVL	SRELKEPPQGAHFLSR	SLDDALKLTEQPE	LANKVDM	WIVGSSVYKEAMNHP	GHLKLFVTRIMQ	
DDFR-EC					MVIGGG	LYLTHI	
					92	97	110 115
	150	160	170	180	186		
DDFR-HM	DFESD	TFPEIDLEKYKLLPEYP	GVLSDVQEEKGIKYK	FEVYEKND			
DDFR-EC							





PCT

特許協力条約に基づいて公開された国際出願

<p>(51) 国際特許分類6 G06F 17/30</p>	<p>A1</p>	<p>(11) 国際公開番号 WO99/62004</p> <p>(43) 国際公開日 1999年12月2日(02.12.99)</p>
<p>(21) 国際出願番号 PCT/JP98/02302</p> <p>(22) 国際出願日 1998年5月26日(26.05.98)</p> <p>(71) 出願人 (米国を除くすべての指定国について) 株式会社 医薬分子設計研究所(INSTITUTE OF MEDICINAL MOLECULAR DESIGN, INC.)[JP/JP] 〒113-0033 東京都文京区本郷5丁目24番5号 角川本郷ビル4F Tokyo, (JP)</p> <p>(72) 発明者 ; および (75) 発明者 / 出願人 (米国についてののみ) 板井昭子(ITAI, Akiko)[JP/JP] 板井玲子(ITAI, Reiko)[JP/JP] 〒113-0033 東京都文京区本郷5-16-6 Tokyo, (JP) 富岡伸夫(TOMIOKA, Nobuo)[JP/JP] 〒113-0033 東京都文京区本郷5-21-12 マンションユキ302号 Tokyo, (JP) 今村正純(IMAMURA, Masazumi)[JP/JP] 〒267-0066 千葉県千葉市緑区あすみが丘4丁目39番地 ガーデンコート社の街五番館301号 Chiba, (JP)</p>		<p>(74) 代理人 弁理士 今村正純, 外(IMAMURA, Masazumi et al.) 〒104-0031 東京都中央区京橋1丁目5番5号 KRFビル5階 Tokyo, (JP)</p> <p>(81) 指定国 AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, KE, KG, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, 欧州特許 (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI特許 (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO特許 (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), ユーラシア特許 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM)</p> <p>添付公開書類 国際調査報告書</p>
<p>(54)Title: METHOD OF ESTIMATING PROTEIN FUNCTIONS</p> <p>(54)発明の名称 蛋白質の機能推定法</p> <pre> DRFR-BC 10 20 30 40 50 60 70 80 90 100 N1SLIAALVYVIGNEKHPHMLPAOLAWETLIDPFIINGNTTCE3ICAPLPQAKM1155QFOTDORVHVWKEBA1AACGOVPEHWVIGGCNT 1122211 111221122111111 11111121221212112 110 120 130 140 150 160 EQFLPKAOKLYLTHIDAEVGOETHFOTEPDOWESVFEFRDAGQMSBVCFRILEMA 1112111  TRTP 10 20 30 40 50 60 70 80 90 100 1VGGYTCCANTYFQVSLMSGYHFCOSLINSQVYBAANTKSGIQVRLGCEDMVWVECHQFESABES1VHPSTYFANTLMDLWILKBAASLMSAV 111111 111111  AS1SLPTKCACTQCL18GNGYHESAGTSTPDLKCLAP1LEOBCKSATPCOITERNFCAGTLEGGDRSCGDRGQPVCSKRLQOIVSMGSGCAQR 111111111 111111 11122221121111 11222222211  210 220 NKPQVTRVCHYVSMIRQT1ASH 1222211  RNAS 10 20 30 40 50 60 70 80 90 100 KETAAAREPQMDSTEAASSSHYCHQNHNSNLTIDKPCFYHTFVNESLADVOAYCQHWVACRNGQTCYQBYSTAS1TOCBRTGSSKYPHCAFTT 111111111221 2122211 111121  210 220 QANRHS1VACCGMYVVPVHTDABV 11111111 111222211  MTEL 10 20 30 40 50 60 70 80 90 100 VLECEGMOLVYHVAEADYVAGHGGDILRLFKSHPELTKFORHRLATEHMAFSEDLKRMGYTLALGAILKARGHNEASLHPLAQENATKSHIP 111111 111221221211 1111112221 12121  210 220 230 IKYLEF18GA11VYLSNRPDGFADAGQAMHALLFLPND1AAKTKELGTQC 11222121111 1111211111111 </pre> <p>(57) Abstract</p> <p>A database containing information on the amino acid sequences of proteins of which one or more biological functions are known and further containing information on each of the amino acid residues constituting each of the above amino acid sequences together with additional information on the score of importance of each amino acid residue in exhibiting the known biological functions; and a method of preparing an alignment which represents the homology among various sites each having a high score of importance in exhibiting the biological functions by estimating the proteins contained in the above database and polypeptides of which biological functions are unknown as yet with respect to the homology between them relative to the agreement between the constituent amino acids while taking into consideration the score of importance in exhibiting the biological functions.</p>		

## (57)要約

1 又は 2 以上の生物学的機能が知られている蛋白質のアミノ酸配列の情報を含み、該アミノ酸配列を構成する各アミノ酸残基について該生物学的機能の発現に関する重要度のスコアを付加した情報を含むデータベース、及び該データベースに格納された蛋白質及び生物学的機能が未知のポリペプチドについて、それぞれの構成アミノ酸の一致に対して生物学的機能の発現に関する重要度のスコアを考慮した相同性の評価値を求め、該重要度の高い部位の相同性を表わしたアラインメントを作成する方法。

PCTに基づいて公開される国際出願のパンフレット第一頁に掲載されたPCT加盟国を同定するために使用されるコード(参考情報)

AE	アラブ首長国連邦	DM	ドミニカ	KZ	カザフスタン	RU	ロシア
AL	アルバニア	EE	エストニア	LC	セントルシア	SD	スーダン
AM	アルメニア	ES	スペイン	LI	リヒテンシュタイン	SE	スウェーデン
AT	オーストリア	FI	フィンランド	LK	スリランカ	SG	シンガポール
AU	オーストラリア	FR	フランス	LR	リベリア	SI	スロベニア
AZ	アゼルバイジャン	GA	ガボン	LS	レソト	SK	スロヴァキア
BA	ボスニア・ヘルツェゴビナ	GB	英国	LT	リトアニア	SL	シエラレオネ
BB	バルバドス	GD	グレナダ	LU	ルクセンブルグ	SN	セネガル
BE	ベルギー	GE	グルジア	LV	ラトヴィア	SZ	スワジランド
BF	ブルキナファソ	GH	ガーナ	MA	モロッコ	TD	チャド
BG	ブルガリア	GM	ガンビア	MC	モナコ	TG	トーゴ
BJ	ベナン	GN	ギニア	MD	モルドヴァ	TJ	タジキスタン
BR	ブラジル	GW	ギニア・ビサウ	MG	マダガスカル	TZ	タンザニア
BY	ベラルーシ	GR	ギリシャ	MK	マケドニア旧ユーゴスラヴィア共和国	TM	トルクメニスタン
CA	カナダ	HR	クロアチア	ML	マリ	TR	トルコ
CF	中央アフリカ	HU	ハンガリー	MN	モンゴル	TT	トリニダード・トバゴ
CG	コンゴ	ID	インドネシア	MR	モリタニア	UG	ウガンダ
CH	スイス	IE	アイルランド	MW	マラウイ	US	米国
CI	コートジボワール	IL	イスラエル	MX	メキシコ	UZ	ウズベキスタン
CN	中国	IN	インド	NE	ニジェール	VN	ヴェトナム
CM	カメルーン	IS	アイスランド	NG	ナイジェリア	YU	ユーゴスラビア
CN	中国	IT	イタリア	NL	オランダ	ZA	南アフリカ共和国
CR	コスタリカ	JP	日本	NO	ノルウェー	ZW	ジンバブエ
CU	キューバ	KE	ケニア	NZ	ニュージーランド		
CY	キプロス	KG	キルギスタン	PL	ポーランド		
CZ	チェコ	KP	北朝鮮	PT	ポルトガル		
DE	ドイツ	KR	韓国	RO	ルーマニア		
DK	デンマーク						

## 明 細 書

## 蛋白質の機能推定法

## 技術分野

本発明はアミノ酸配列の情報を基にして、そのアミノ酸配列から構成される蛋白質の機能を推定する方法に関するものである。より詳しく言うと、アミノ酸配列の情報を基にして、そのアミノ酸配列から構成される蛋白質の酵素活性などの生物学的機能をコンピュータで利用可能な特定のデータベースを利用して効率的に推定する方法に関する。

## 背景技術

蛋白質は生体内で生命活動の維持に必須の物質であり、動物や植物に限らず微生物にもさまざまな蛋白質が存在しており、それぞれ固有の機能や役割を担っている。蛋白質をその機能から大別すると、化学反応を触媒する酵素、シグナル伝達物質の受け皿蛋白である受容体、それ自身がシグナルを伝えるシグナル伝達蛋白、及び特定の物質を結合して輸送する蛋白などに分類でき、それぞれが多様な機能によりさらに細分される。例えば、特定の基質の特定部位を還元する酵素や、蛋白質を加水分解する酵素などのように、酵素はそれぞれ固有の反応を触媒している。

蛋白質は主として 20 種のアミノ酸によって構成されており、50～1000 個程度のアミノ酸がさまざまな順序でペプチド結合により鎖状につながったポリペプチド分子である。蛋白質ごとにアミノ酸のつながる順序（アミノ酸配列または 1 次構造と呼ぶ）が異なっており、その結果、それぞれの蛋白質は異なる生理機能を発現できるようになる。すなわち、長いポリペプチド鎖が折り畳まれて立体構造が形成されると、標的分子（酵素基質分子や受容体基質分子など）の捕捉が可能になり、反応に関わる官能基が適切な位置関係に配置されるなど、目的の生物学

的作用の発現に適した場が提供される。それぞれのアミノ酸配列から固有の立体構造が決まり、その立体構造から生物学的機能が決定されることは容易に推定されるが、それらの関係の必然性は未だうまく説明されていない。

蛋白質の研究は、酵素活性を指標として蛋白質を単離・精製し、分子量や構成アミノ酸数、アミノ酸毎の数を決定した後にアミノ酸配列を決定するという古典的な方法に代わって、末端の 20 残基ほどを決定して対応する遺伝子配列を合成した後、蛋白質をコードする遺伝子を釣り上げて遺伝子側から全部のアミノ酸配列が決定する方法が利用されるようになった。これらの研究ではすでに機能がわかっている蛋白質を対象としてきたが、最近では全く逆の順序で研究が行われることも多い。その理由は、遺伝子配列の解析が極めて容易になったために、蛋白質を単離することなく、遺伝子の側からそれがコードするはずの蛋白質のアミノ酸配列を容易に決定できるようになったことにある。

その結果として、生物学的機能がわからないままアミノ酸配列だけが推定できる蛋白質が急増している。蛋白質の生物学的機能は立体構造に基づいて発現されるので、このような機能未知の蛋白質の立体構造を結晶解析や n m r 解析で解析して、その生物学的作用を推定する試みがなされている。しかしながら、このような構造解析には生化学的な使用量よりはるかに大量で、かつ高純度の試料が必要である。また、立体構造からただちに生物学的機能が推定されるわけではなく、仮に生物学的機能が推定されたとしてもそれが重要なものであるとは限らないので、研究の投資効率が極端に悪いという問題がある。したがって、蛋白質の立体構造を決定する前に、そのアミノ酸配列を有する蛋白質の生物学的作用を推定する方法の開発が切望されている。このような方法が開発されれば、蛋白質研究や遺伝子研究に多大な貢献があるものと期待される。

立体構造と生物学的機能は密接な関係をもっており、機能が既知の蛋白質の立体構造情報は機能メカニズムを説明するだけでなく、さまざまな目的に役に立つ。蛋白質またはリガンド分子との複合体の 3 次元座標はプロテインデータバンク (Brookhaven National Laboratories, U. S. A.) に収められており、世界中で利

用できるようになっている。現在、収録されている構造数は約 5000 程度であるが、生物種の違いやミュータントを除いた独立の蛋白質で考えると約 400~500 程度である。解析技術の普及と蛋白質の単離精製の技術の進歩などから、結晶解析される蛋白質の数は加速度的に増加しつつあるが、現状では立体構造が解明されないままの蛋白質が圧倒的に多い。

結晶解析や n m r 解析によらずに蛋白質の立体構造やリガンド分子との相互作用の様子を推定する方法としてモデリングを利用することができる。アミノ酸配列においてある程度相同性の高い類似蛋白質の立体構造が既に解析されている場合には、その構造を鋳型として利用してモデリングを行うことにより、アミノ酸残基の対応関係に基づいた立体構造を構築できる。この方法は試料を入手する必要がなく、一般にはコンピュータグラフィックス画面上で対話的に行える点で優れた方法である。例えば、一致していないアミノ酸については側鎖を置換することによって行われる。側鎖のコンフォメーションの問題や挿入または欠損アミノ酸の主鎖の問題を有しているものの、推定構造の信頼性はアミノ酸配列の相同性の高さによって決定され、結晶構造とほぼ同様の扱いが可能である。

アミノ酸の種類ができるだけ一致するように 2 種以上の蛋白質のアミノ酸配列間の対応関係をつける方法（アラインメント）は、モデリングの目的だけでなく、生物種間やファミリー間での類似性や相違点を調べる目的で頻繁に利用されている。アラインメントの手法では、概念的には一方の配列を他方に対して 1 残基ずつずらしながらアミノ酸の一致のスコアが最もよい対応位置を見つけることになるが、実際には、配列間の対応関係の可能性は無限にあるのできめ細かな配慮と繰り返し操作が必要になり、精度の高い結果を得るためには極めて煩雑な作業が必要になる。例えば、一方の配列に挿入や欠損がある場合も多いので単純に全配列でのアミノ酸の一致度をスコアにすることはできず、部分的に良く一致する部分配列を探すことが必要であり、所定の残基数単位（ウィンドウ）で一致のスコアを算出する必要もある。場合によっては、完全な一致ではなく性質の似たアミノ酸を相同としてスコアを求める必要もある。

一般に、相同性が低い場合にはアラインメントも一義的には決まらず曖昧さが残るという問題を有しているものの、機能が既知の蛋白質群からアミノ酸配列上の類似性の高い蛋白質を探し出すことができるので、アラインメントは現在のところアミノ酸配列しか手がかりのない蛋白質についてその機能を推定する最も簡便な方法である。このような理由から、アラインメントの煩雑な操作をコンピューターを用いて部分的に自動化する試みがなされている。例えば、アミノ酸配列のマッチング法としてFASTA (Pearson, W. R. and Lipman, D. J., Proc. Natl. Acad. Sci. USA, 85, pp. 2444-2448, 1988) と BLAST (Altschul, S. F. et al., J. Mol. Biol., 215, pp. 403-410, 1990) が知られている。これらの方法は、短い特定の配列が長い対象配列中に存在するか否かを高速に調べるのに適しているものの、長い配列同士を比較する場合や、相同性が低く断片的にしか一致していないような配列を検索対象とする場合には、類似性の判断や類似部分の抽出が著しく困難であり、相同性の判定精度が低い。従って、これらの方法はアラインメントや蛋白質の機能推定の目的には不十分であり、さらに精度に優れた高速な方法の開発が求められていた。

#### 発明の開示

本発明の課題は、アミノ酸配列の情報を基にして、そのアミノ酸配列から構成される蛋白質の生物学的機能を推定する方法を提供することにある。さらに詳しくは、アミノ酸配列の情報のみが利用可能である場合に、生物学的機能がすでに知られている蛋白質のアミノ酸配列との相同性を特定のデータベースを利用してコンピューターにより効率的に検索し、該アミノ酸配列から構成される蛋白質の生物学的機能を正確かつ高速に検索する方法を提供することにある。

本発明者らは上記の課題を解決すべく鋭意努力した結果、下記の特徴を有するデータベースを用いると、アミノ酸配列から蛋白質の機能を極めて高速かつ高精度に推定できることを見出した。

すなわち本発明は、1又は2以上の生物学的機能が知られている蛋白質のアミ

ノ酸配列の情報を含み、該アミノ酸配列を構成する各アミノ酸残基について該生物学的機能の発現に関する重要度のスコアを付加した情報を含むデータベースを提供するものである。

このデータベースは、例えば、アミノ酸配列の相同性に基づいて生物学的機能が未知な蛋白質の機能を推定するために用いることができ、好ましい態様では、生物学的機能が知られている蛋白質のアミノ酸配列の情報として、蛋白質の3次元構造などの立体構造に関する情報が利用可能な蛋白質のアミノ酸配列を用いて、該アミノ酸配列を構成する各アミノ酸残基について該蛋白質とリガンド分子との結合や生物学的機能の発現に関する重要度のスコアが付加されている。これらのデータベースは、一般的には、フロッピーディスク、CD-ROM、磁気テープ、光ディスクなどの種々の記憶用媒体に格納することが可能である。

別の観点からは、上記データベース中の蛋白質（本明細書において「鑄型蛋白質」と呼ぶ場合がある）と生物学的機能が未知のポリペプチド（本明細書において「対象蛋白質」と呼ぶ場合がある）について、それぞれの構成アミノ酸の一致に対して生物学的機能の発現に関する重要度のスコアを考慮した相同性の評価値を求め、該重要度の高い部位の相同性を表わしたアラインメントを作成する方法が本発明により提供される。

この方法の好ましい態様では、生物学的機能の発現に関して重要度が高い連続した2以上のアミノ酸残基を含むグループ配列を用いて、上記データベース中の蛋白質と対象蛋白質について相同性の高い対応関係を検索する工程を含んでいる。また、他の好ましい態様では、データベース中の1の蛋白質と対象蛋白質について上記アラインメントから相同性の最終評価値を得る工程と、データベースに含まれる全蛋白質についての最終評価値から生物学的機能に関して対象蛋白質に最も類似した蛋白質を推定する工程を含んでいる。これらの方法では、蛋白質全体としての相同性は低くても、生物学的機能に関係した部位の相同性が高い蛋白質を効率よく抽出することができ、対象蛋白質の機能を高速かつ高精度に推定できるという特徴を有している。

### 図面の簡単な説明

第1図は、生物学的機能と立体構造がわかっている4種類の蛋白質のアミノ酸配列を構成する各アミノ酸残基について、それぞれの生物学的機能の発現に関する重要度のスコアを付加したアミノ酸配列の情報を示す図である。図中の記号は、それぞれ、大腸菌由来のジヒドロ葉酸還元酵素 (DHFR-EC)、ウシ由来のトリプシン (TRYP)、ウシ由来のリボヌクレアーゼA (RNAS)、クジラ由来のミオグロビン (MYGL) を示し、アミノ酸残基は1文字表記で示した。

第2図は、対象蛋白質(DHFR-HM)と、該対象蛋白質に対して最も高い評価値SSを与える鋳型蛋白質として抽出されたDHFR-ECとのアラインメントを示した図である。図中の記号は、それぞれ、ヒト由来のジヒドロ葉酸還元酵素(DHFR-HM)及び大腸菌由来のジヒドロ葉酸還元酵素 (DHFR-EC)を示し、1段目はDHFR-HMのアミノ酸番号、2段目はDHFR-HMのアミノ酸配列、3段目はDHFR-ECのアミノ酸配列の部分配列、4段目はDHFR-ECのアミノ酸配列の部分配列のアミノ酸番号を示す。

### 発明を実施するための最良の形態

本発明のデータベースは、1又は2以上の生物学的機能が知られている蛋白質のアミノ酸配列の情報を含み、該アミノ酸配列を構成する各アミノ酸残基について該生物学的機能の発現に関する重要度のスコアを付加した情報を含むことを特徴としている。格納の対象となる蛋白質は、例えば、酵素作用や受容体作用などの1又は2以上の生物学的機能が知られており、全アミノ酸配列が知られているものであればいかなるものでもよいが、好ましくは、その蛋白質の立体構造やリガンド結合部位の立体構造がすでに解明若しくは推定されており、または容易に推定可能なものであることが好ましい。データベースには立体構造が明らかにされた蛋白質に関する情報になるべく多く含まれていることが望ましい。例えば、蛋白質またはリガンド分子との複合体の3次元座標はプロテインデータバンク



(Brookhaven National Laboratories, U.S.A.) に収められているが、約 5000 程度(生物種の違いやミュータントを除いた独立の蛋白質で考えると約 400~500 程度)の蛋白質についての情報が利用可能であり、本発明のデータベースの作成に好適に用いることができる。

本発明のデータベースでは、生物学的機能が知られた蛋白質を構成するアミノ酸配列を基にして、各々のアミノ酸残基について該生物学的機能の発現に関する重要度のスコアを付加することを特徴としている。それぞれの蛋白質について格納する情報としては、例えば、蛋白質名、生物種、臓器・器官、サブタイプ、機能の種類、機能の細分類(例えば、酵素については蛋白分解作用や還元作用などの酵素作用)、酵素の分類番号(EC 番号)、酵素反応や生物学的機能に関係するリガンド分子(酵素基質、受容体基質、補酵素、金属イオンなど)、立体構造の由来(例えば、X線結晶解析、n m r 解析、同様な生物学的機能を有する類似蛋白質の情報に基づくモデリングによるものなど)、主たる引用文献名、他のデータベースの参照番号、対象部位リガンド、及び全アミノ酸配列などを挙げるができるが、これらに限定されることはなく、適宜の情報を追加又は削除してもよい。

本発明のデータベースは、上記の情報に加えて、該アミノ酸配列を構成する各アミノ酸残基について該生物学的機能の発現に関する重要度のスコアを付加した情報を含んでいる。重要度のスコアは、例えば、重要度が全くないものについてはゼロ、重要度が極めて高いものについては 10 などの数字やその他の記号を与えることにより行われるが、好ましくは、重要度に関連する 2 以上要素を勘案しながら算出するのが一般的である。本発明のデータベースには、さらに、生物学的機能の発現に寄与がある(スコアがゼロではない)連続したアミノ酸残基配列(1~n)の有無、スコアづけの方式、スコアの合計、蛋白質間でスコアの合計などを規格化のためのスケール因子などの情報を付加することができるが、付加すべき情報はこれらに限定されることはなく、適宜の情報を追加又は削除してもさしつかえない。なお、一個の蛋白質に複数の生物学的機能や複数のリガンド分

子が知られている場合には、それぞれについての情報を格納しておくことが好ましい。

以下、蛋白質のアミノ酸配列を構成する各アミノ酸残基について該生物学的機能の発現に関する重要度のスコアを与える具体的な手法を説明するが、これらの説明は単に例示のためにのみ示されたものであり、いかなる意味においても限定的に解釈してはならない。また、本発明のデータベースはこれらの手段によって製造されたものに限定されることはない。なお、以下の説明においては、重要度のスコアを数値で表わす例について説明しており、重要度が全くないものについてはゼロ、重要度が高くなるにつれて大きな数値を与えるようにしているが、スコア付けの方法がこのような方法に限定されないことも理解すべきである。

(a) 低分子量のリガンド分子を含む蛋白質複合体の結晶解析がなされている場合：

プロテインデータバンクに収納された蛋白質であって、酵素基質、受容体基質、または阻害剤などの低分子量のリガンド分子との複合体の3次元構造が解析されている場合には、リガンド分子からの各アミノ酸残基の距離を求めて、各アミノ酸残基に距離に応じた重要度のスコアを与えることができる。低分子量のリガンド分子としては、例えば、薬理学的に活性な有機化合物、酵素基質、金属イオンなど、いかなるものであってもよい。例えば、リガンド分子のいずれかの原子（例えば、リガンドとなるCa原子やリガンド分子中の側鎖の1原子など）から10Å以内にあるアミノ酸残基には1、8Å以内なら2、6Å以内なら4、10Åより大きい場合には0などのスコアを与えることが可能である。

(b) 蛋白質単独で結晶解析がなされている場合：

リガンド分子を含まずに結晶解析されている蛋白質についても、さまざまな実験から生物学的機能（酵素活性など）に関係する構造領域が推定される場合には、その近傍のアミノ酸残基に対して上記(a)で行ったような距離に応じた数値を与えることが可能である。生物学的機能と立体構造との対応がついていない場合においても、生物学的機能が明確である場合には、その蛋白質の立体構造をコンピ

ュータグラフィックス画面上に描いて回転させながら顕著な内孔を探ることによって、生物学的機能に関するアミノ酸残基を抽出することができる。

(c) 構造保存領域：

同じ生物学的機能を有する蛋白質について、アミノ酸配列の異なるサブタイプや異生物種の蛋白質など2種以上の立体構造の解析結果を利用できる場合には、それらの構造の重ね合わせを行うことにより、立体構造的に保存された領域を抽出することができ、それらの領域に含まれるアミノ酸残基に高い数値を与えることができる。

(d) モデリング：

蛋白質の立体構造が解析されていない場合、実質的に同一の生物学的作用を有することが知られている類縁蛋白質の立体構造を基にして構築したモデリング構造に基づいて、重要度のスコアをつけることが可能である。例えば、受容体サブタイプやアイソザイム、同じファミリーに属する蛋白質、異生物種の同一機能の蛋白質であってアミノ酸配列の相同性が高い場合にはモデリング構造の信頼性が高いことが知られている。重要度のスコアの与え方は、例えば、上記の各手法と同様に行えばよい。

(e) 生化学実験や遺伝子実験：

例えば生化学的な実験などから生物学的機能の発現に重要であることが推定されるアミノ酸残基や、遺伝子的なアミノ酸の変換の実験（ポイントミューテーションなど）から酵素作用などの生物学的機能の発現に必須であると推定されるアミノ酸残基については、高い重要度を与えることができる。例えば、酵素反応においては、リガンド分子との結合などの観点での評価に加えて、触媒的な役割を果たすアミノ酸残基に大きな数値を与えることが可能である。

(f) 高分子リガンド分子である蛋白質：

一般的に、低分子量のリガンド分子との結合が機能に必須である蛋白質はそのリガンド分子と安定に結合するための内孔を有している。一方、例えば蛋白質などの高分子量のリガンド分子が結合する蛋白質では、顕著な内孔をもたずに受容体

蛋白質と分子表面で結合することが多く、受容体蛋白質も顕著な内孔を有しないことがある。例えば、サイトカインのようにそれ自身が高分子リガンドとなる場合には、モノクローナル抗体を用いて推定されるエピトープ領域のアミノ酸残基に大きな数値を与えてもよい。

以上に例示したような手法の1種又は2種以上の組み合わせを用い、さらに必要に応じて適宜の手法を追加することによって、生物学的機能が知られている蛋白質のアミノ酸配列を構成する各アミノ酸残基について、該生物学的機能の発現に関する重要度のスコア付けを行い、重要度のスコアが付加されたアミノ酸配列の情報を作成することができる。なお、アミノ酸残基と生物学的機能の発現との関係については、例えば上記(a)の手法などによりその関係が十分に実証されているものに加えて、その関係がある程度推定可能なものなど、種々の基準をスコア付けに利用できることは言うまでもない。

例えば、結晶構造が既知の蛋白質や、生物学的機能の面から類似の立体構造を有することが推定される蛋白質など、なるべく多くの蛋白質について重要度のスコアが付加されたアミノ酸配列の情報を収集し、コンピューターが利用可能な所定の形式で格納して本発明のデータベースを構築することができる。この際、それぞれの蛋白質の情報の量と質に応じて、それぞれの蛋白質について適宜の異なる基準によりスコアを付加してもよい。もっとも、データベース中にスコアづけの方式と蛋白質間の合計スコアの規格化のためのスケール因子を加えることが必要になる場合もある。なお、上記のような情報の入力は、一定の方式に従ってマニュアルで行うことも可能であるが、一般的には、コンピュータグラフィックス画面上で所定のプログラムを用いて行うのが効率的である。

本発明の方法では、上記データベースを用いて、データベース中に情報が格納された鋳型蛋白質と生物学的機能が未知の対象蛋白質について、アミノ酸残基の重要度のスコアから計算した相同性の評価値が最大になるようにアラインメントを作成し、ついで、データベースに含まれる2以上の鋳型蛋白質、好ましくは全部の鋳型蛋白質について同様なアラインメントを作成した後、鋳型蛋白質間で相同

性の評価値の比較を行い、評価値が最も高い鋳型蛋白質を選出することができる。このようにして選出された鋳型蛋白質は、対象蛋白質との立体構造の類似度が高く、実質的に同一の生物学的機能を有する蛋白質であると推定することが可能である。

上記の方法は、一般的には、本発明の上記データベース中の鋳型蛋白質に関する情報を1つ1つ取り出して、対象蛋白質のアミノ酸配列に対してアラインメント作業をすることにより行われる。対象蛋白質のアミノ酸配列の情報が直接利用できる場合にはその情報を入力して用いればよく、対象蛋白質をコードする遺伝子配列の情報のみが利用可能である場合には、その核酸配列の情報から対象蛋白質のアミノ酸配列の情報を翻訳して用いる必要がある。

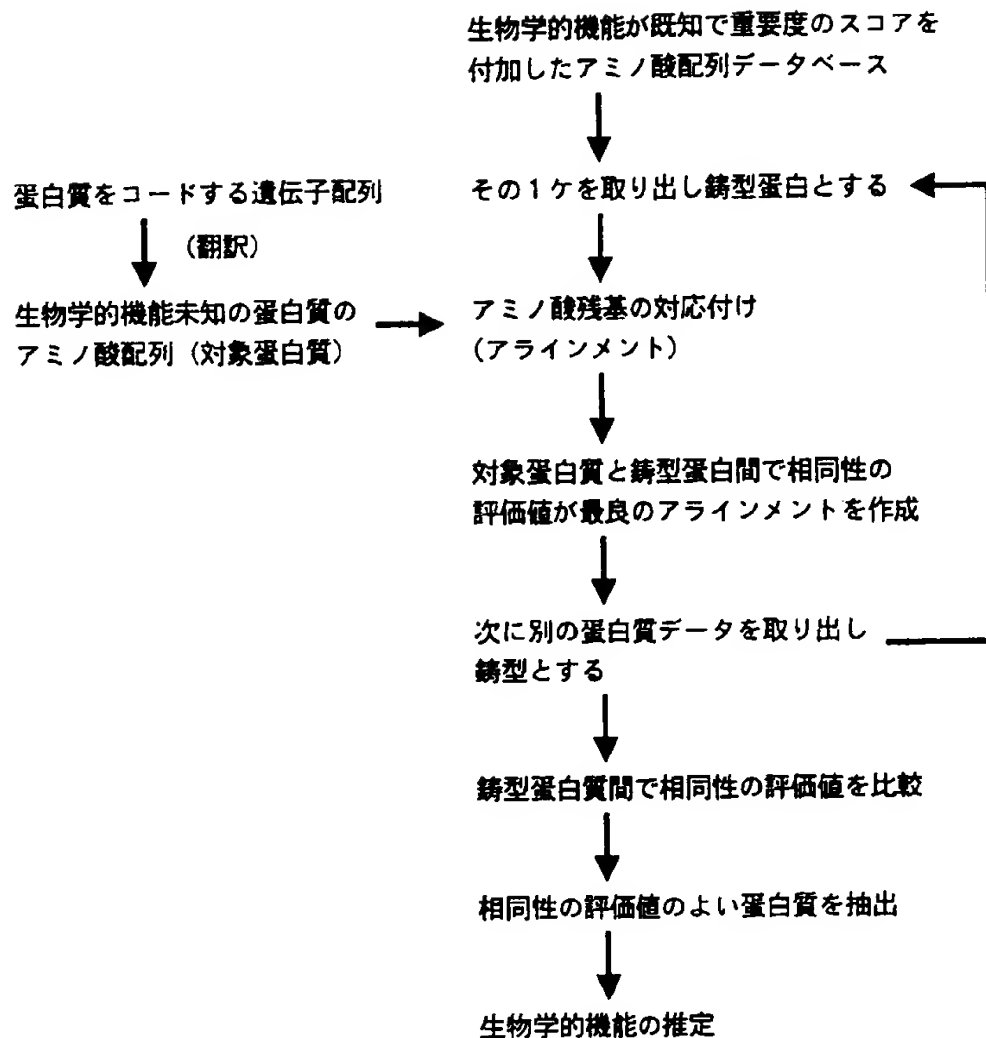
本発明の好ましい方法の一例として、鋳型蛋白質のアミノ酸配列において生物学的機能への寄与がある連続した2以上のアミノ酸残基（スコアがゼロでない連続したアミノ酸残基）を含むグループ配列を用いて、鋳型蛋白質と対象蛋白質について相同性の高い対応関係を検索する方法を挙げることができる。もっとも、アラインメント作業はこの方法に限定されることはなく、当業者に利用可能ないかなる方法で行ってもよい。

グループ配列を利用する上記の方法では、グループ配列を対象蛋白質のアミノ酸配列に対して1残基ずつずらしながら、グループ毎に相同性の評価値を求めることができ、その後、必要に応じてグループ配列の繋がる順序や長さなどの因子を考慮しつつ、全グループ配列の合計評価値が最良になるように、対象蛋白質のアミノ酸配列に対する各グループ配列の対応関係を決定することができる。この手順をデータベース中のすべての鋳型蛋白質について行い、高い総スコアを有する1又は2個以上の蛋白質を抽出することができる。対象蛋白質は、このようにして抽出された鋳型蛋白質と実質的に同一の生物学的機能を有している可能性が高い。

相同性の評価値としては、例えば、鋳型蛋白質中のグループ配列と対象蛋白質配列中の対応アミノ酸残基が一致した場合には、一致したアミノ酸残基に重要度の

スコアを転記して単に合計するのが簡単である。もっとも、重要度のスコアの高いアミノ酸の一致を重視したアラインメントを作成するためには、重要度の各スコアをさらに1又は2以上の関数で処理して用いてもよい。対象蛋白質とデータベース中に含まれる全鋳型蛋白質のアラインメントを作成するにあたっては、異なるアラインメント間で全相同性評価値を比較すればよいが、一般的には、大きさの異なる鋳型蛋白質間においても重要度のスコアの合計値によって対象蛋白質との相同性の良し悪しを比較できるように、重要度のスコアの規格化のためのスケール因子を鋳型蛋白質毎に算出してデータベースに格納しておくことが望ましい。対象蛋白質と各鋳型蛋白質とのアラインメント作業が完了して相同性のスコアが計算された段階で、各鋳型蛋白質の相同性スコアに対応のスケール因子を掛け合わせて最終スコアを求め、鋳型蛋白質間での相同性の優劣を決定することができる。。

また、異なる生物種に存在する同種蛋白質において、例えば、側鎖の長さが炭素原子1個分違うだけでカルボキシル基を共通に有するアスパラギン酸とグルタミン酸とが、アミノ酸配列中の同じような位置で同一の役割を果たしていることがある。このような場合には、アミノ酸残基の一致を判定するに当たり、これらのアミノ酸残基を一致しているとみなすのが妥当である。また、ロイシンとイソロイシン、バリンなどのアミノ酸残基は、形状やサイズ（嵩高さ）では異なるものの、疎水性の観点からは類似の性質を有している。従って、アミノ酸配列の相同性を数値化する際には、このような類似アミノ酸残基の存在が反映されるように、アミノ酸残基の類似度を段階化した対応表を用いることが望ましい。このようなアミノ酸残基の類似度としてはいかなるものを用いてもよいが、類似度を記載した対応表として、例えば、PAM250 (Dayhoff, M. O., et al., Atlas of Protein Sequence and Structure, Dayhoff, M. O. Ed., Vol. 5, Suppl. 3, pp. 345-352, NBRF, Washington, 1978) や BLOSUM (Henikoff, S. and Henikoff, J. G., Proc. Natl. Acad. Sci. USA, 89, pp. 10915-10919, 1992) などが利用可能である。



本発明の方法の一例を概念図として上記に示した。また、本発明の方法は、例えば下記の工程を含む方法であってもよい。もっとも、本発明の方法はこれらの方法に限定されることはなく、これらの方法において採用された工程に加えて、必要に応じて1又は2以上の適宜の工程を追加することができ、所望であれば1又は2以上の工程を省略できる場合があることを理解すべきである。このような修飾ないし改変された方法がすべて本発明の範囲に包含されることは言うまでもない。

(1) 対象蛋白質のアミノ酸配列を呼び出す工程；

- (2) 上記データベースから鑄型配列を1個取り出す工程；
- (3) 鑄型蛋白質のアミノ酸配列から重要度のスコアが一定値以上の部分配列 a, b, c, d, e, ---, n, ---- を例えばN末端から順に取り出す工程（それぞれの部分配列の長さを  $1a, 1b, 1c, 1d, 1e, \dots, 1n, \dots$  とする）；
- (4) 部分配列 a を対象配列の1番目に位置付け、1アミノ酸残基ずつずらしながら相同性の評価値  $S(a)_i$  を算出する工程（相同性の評価値として一致したアミノ酸残基の重要度のスコアを加える）；
- (5) 部分配列 b を対象配列の  $(1+1a)$  番目に位置づけ、1アミノ酸残基ずつずらしながら、相同性の評価値  $S(b)_i$  を算出する工程（相同性の評価値として一致したアミノ酸残基の重要度のスコアを加える）；
- (6) 同様に c, d, e, ---, n, ---- について相同性の評価値  $S(n)_i$  を算出する工程；
- (7) a, b, c, d, e, ---, n, ---- の順序とそれぞれのアミノ酸残基数を考慮しながら、全部分配列の相同性  $SS$  がもっとも大きくなるように部分配列の対応位置を決定する工程；
- (8)  $SS$  にスケール因子をかけて  $SSS$  とする工程；
- (9) データベース中の全鑄型蛋白質について上記の手順を行い  $SSS$  を求める工程；及び
- (10)  $SSS$  の高い蛋白質を抽出する工程

#### 実施例

本発明を下記の実施例により更に具体的に説明するが、本発明の範囲は下記の実施例に限定されることはない。

#### 例1：データベースの作成

生物学的機能と立体構造がわかっている大腸菌由来のジヒドロ葉酸還元酵素 (DHFR-EC)、ウシ由来のトリプシン (TRYP)、ウシ由来のリボヌクレアーゼ A



(RNAS)、クジラ由来のミオグロビン (MYGL) の4種類の蛋白質からなるデータベースを作成した。それぞれの結晶構造はプロテインデータベース (Brookhaven National Laboratories, U.S.A.) から入手した。アミノ酸残基のいずれかの構成原子が、蛋白質に結合しているリガンド分子 (阻害剤又は補酵素) のいずれかの原子から 4 Å 以内にある場合には2、4~10 Å の範囲内である場合には1、その他の場合には0を与えて、それぞれのアミノ酸配列を構成する各アミノ酸残基についてそれぞれの生物学的機能の発現に関する重要度のスコアを付加したアミノ酸配列の情報を作成した。図1に各蛋白質のアミノ酸配列とスコア付けの結果を示す。

#### 例2：生物学的機能の推定

ヒト由来のジヒドロ葉酸還元酵素 (DHFR-HM) を対象蛋白質とし、上記データベースを用いて、該対象蛋白質の生物学的機能を本発明の方法により推定した。DHFR-HM は生物学的機能及び立体構造が知られている蛋白質であるが、生物学的機能及び立体構造を未知なものとして解析を行った。データベース中の鋳型蛋白質のそれぞれのアミノ酸配列について、スコア値が1以上の部分配列を取り出し、対象蛋白質のアミノ酸配列に対して1残基ずつずらしながら相同性の評価値Sを算出し、評価値Sの最も高いアラインメント位置を決定した。

評価値Sの算出には、アミノ酸の類似度に関する対応表 BLOSUM62 を用い、部分配列と対象蛋白質のアミノ酸配列とのアミノ酸残基の組に応じた類似度の評価値を表から求めて、部分配列の各残基のスコアと類似度の評価値との積を部分配列の長さによって積算する方法を採用した。全部分配列の相同性SSは、部分配列ごとに評価値Sの最大値を求めて、それらの和として算出した。相同性決定に用いた配列の長さの違いを補正するために、スケール因子として全部分配列のスコアの総和の逆数を用いて、スケール因子をSSの値に乗じて最終的な評価値SSSを算出した。この結果、表1に示すように、DHFR-EC, TRYP, RNAS, MYGL を鋳型蛋白質として用いた場合、DHFR-EC が最も高い評価値SSSを与え、対象蛋

白質(DHFR-HM)が DHFR-EC と類似の蛋白質であり、ジヒドロ葉酸還元酵素の活性を有するものと推定された。図2には DHFR-HM と DHFR-EC のアラインメントを示した。

表 2

蛋白質	評価値 SSS
DHFR-EC	1.82
TRYP	1.09
RNAS	1.22
MYGL	0.61

#### 産業上の利用可能性

本発明のデータベースは、アミノ酸配列の情報を基にして、そのアミノ酸配列から構成される蛋白質の生物学的機能を推定するために有用であり、本発明の方法は、このデータベースを利用してアミノ酸配列から構成される蛋白質の生物学的機能を正確かつ高速に検索することができるので有用である。

## 請 求 の 範 囲

1. 1又は2以上の生物学的機能が知られている蛋白質のアミノ酸配列の情報を  
含み、該アミノ酸配列を構成する各アミノ酸残基について該生物学的機能の発現  
に関する重要度のスコアを付加した情報を含むデータベース。
2. アミノ酸配列の相同性に基づいて生物学的機能が未知な蛋白質の機能を推定  
するために用いる請求の範囲第1項に記載のデータベース。
3. 生物学的機能が知られている蛋白質のアミノ酸配列の情報として、蛋白質の  
立体構造に関する情報が利用可能な蛋白質のアミノ酸配列を用いて作成された請  
求の範囲第1項又は第2項に記載のデータベース。
4. 記憶用媒体に格納された請求の範囲第1項ないし第3項のいずれか1項に記  
載のデータベース。
5. 請求の範囲第1項ないし第3項のいずれか1項に記載のデータベースに格納  
された蛋白質及び生物学的機能が未知のポリペプチドについて、それぞれの構成  
アミノ酸の一致に対して生物学的機能の発現に関する重要度のスコアを考慮した  
相同性の評価値を求め、該重要度の高い部位の相同性を表わしたアラインメント  
を作成する方法。
6. 生物学的機能の発現に関して重要度が高い連続した2以上のアミノ酸残基を  
含むグループ配列を用いて、上記データベース中の蛋白質及び対象蛋白質につい  
て相同性の高い対応関係を検索する工程を含む、請求の範囲第5項に記載の方法。
7. データベース中の1の蛋白質と対象蛋白質について上記アラインメントから  
相同性の最終評価値を得る工程を含む、請求の範囲第5項又は第6項に記載の方  
法。
8. データベースに含まれる全蛋白質についての最終評価値から生物学的機能に  
関して対象蛋白質に最も類似した蛋白質を推定する工程を含む、請求の範囲第7  
項に記載の方法。

**DHFR-EC**

10	20	30	40	50	60	70	80	90	100
MISLIAALAVDRVIGMENAMPWNL	PADLAWFKRNTLDKPVIMGRHTWESI	GRPLPGRKNIILSSQPGTDDRVTWVKS	VDIAAACGDVPEIMVIGGRVY						
112221	1122112211111	11111211221212112							112111

EQFLPKAQKLYLTHIDAEEVGDTHFPDYE

110	120	130	140	150	159
EQFLPKAQKLYLTHIDAEEVGDTHFPDYE	PDWESVFSEFHDADAQN	SHSYCFKILERR			
111211					

**TRYP**

10	20	30	40	50	60	70	80	90	100
IVGGYTCGANTVPYQVSLNSGYHFCG	SLINSQWVVSAAHCYKSGIQVRLGED	NIINNVVEGNEQFISAKSIVHP	SYNSNTLNNDIMLIKLSAASLNSRV						
11111	11111								

ASISLPTSCASAGTQCLISGWGNTKSSG

110	120	130	140	150	160	170	180	190	200
ASISLPTSCASAGTQCLISGWGNTKSSG	TSYDPVLKCLKAPILSDSSCKSA	YPGQITSNMFCAGYLEGGKDS	COGDSGGPVVCSGKLQGI	VSWSGSGCAQK					
1111111111	111111								

NKPGVYTKVCNYSWIKQTIASN

210	220
NKPGVYTKVCNYSWIKQTIASN	
122221	

**RNAS**

10	20	30	40	50	60	70	80	90	100
KETAAAKFERQHMDSSISAASSNYC	NQMMSRNLT	KDRCKPVNTFVHESLADVQAVCS	QKNVACKNGQTCYQSYSTMSITD	CREITGSSKYPNCAYKTT					
11111111221		2122211							

QANKHIIIVACEGNPYVPVHFDA

110	120	124
QANKHIIIVACEGNPYVPVHFDA		
1111111	111222211	

**MYGL**

10	20	30	40	50	60	70	80	90	100
VLSEGEWQLVLHVWAKVEADVAGHG	QDILIRLFKSHPETLEKFD	DRFKHLKTEAEMKASEDL	KKHGVTVLTLGAILKKKGGH	HEAELKPLAQSHATKHKIP					
	111111	111211221211							

IKYLEFISEAIIHVLHSRHPGDFGADAQ

110	120	130	140	150
IKYLEFISEAIIHVLHSRHPGDFGADAQ	GAMNKALELFRKDIAAKYKELGYQG			
11222121111	111121111111			

第1図

## 第2図

DHFR-HM	10	20	30	40	50	60	70
DHFR-EC	3	9	24	38	40		
	VGSLNCIVAVSQNMGIGKNGDLPWPPLRNEFRYFORMTTSSVEGKQNLVIMGKKTWFSIPEKNRPLKGR						
	SLIAALA		LPADLAWFKRNTLDK		VIMGRHTWESIGRPLPGR		
							57
DHFR-HM	80	90	100	110	120	130	140
DHFR-EC							
	INLVLSRELKEPPQGAHFLSRSLDDALKLTEQPELANKVDMMVWIVGSSVYKEAMNHPGHLKLFVTRIMQ						
					MVIGGG		LYLTHI
					92 97		110 115
DHFR-HM	150	160	170	180	186		
DHFR-EC							
	DFESDTFFPEIDLEKYKLLPEYPGVLSDVQEEKGIGIKYKFEVYEKND						

## 第Ⅰ欄 請求の範囲の一部の調査ができないときの意見 (第1ページの1の続き)

法第8条第3項 (PCT 17条(2)(a)) の規定により、この国際調査報告は次の理由により請求の範囲の一部について作成しなかった。

1. ☒ 請求の範囲 1-4 は、この国際調査機関が調査することを要しない対象に係るものである。つまり、  
請求の範囲 1-4 は専らその情報の内容に特徴を有する情報の単なる提示と認められるから、特許協力条約に基づく国際出願等に関する法律施行規則第42条(5)に定められた国際調査を要しない対象である。
2. ☐ 請求の範囲 \_\_\_\_\_ は、有意義な国際調査をすることができる程度まで所定の要件を満たしていない国際出願の部分に係るものである。つまり、
3. ☐ 請求の範囲 \_\_\_\_\_ は、従属請求の範囲であってPCT規則6.4(a)の第2文及び第3文の規定に従って記載されていない。

## 第Ⅱ欄 発明の単一性が欠如しているときの意見 (第1ページの2の続き)

次に述べるようにこの国際出願に二以上の発明があるとこの国際調査機関は認めた。

1. ☐ 出願人が必要な追加調査手数料をすべて期間内に納付したので、この国際調査報告は、すべての調査可能な請求の範囲について作成した。
2. ☐ 追加調査手数料を要求するまでもなく、すべての調査可能な請求の範囲について調査することができたので、追加調査手数料の納付を求めなかった。
3. ☐ 出願人が必要な追加調査手数料を一部のみしか期間内に納付しなかったので、この国際調査報告は、手数料の納付のあった次の請求の範囲のみについて作成した。
4. ☐ 出願人が必要な追加調査手数料を期間内に納付しなかったので、この国際調査報告は、請求の範囲の最初に記載されている発明に係る次の請求の範囲について作成した。

追加調査手数料の異議の申立てに関する注意

- ☐ 追加調査手数料の納付と共に出願人から異議申立てがあった。  
☐ 追加調査手数料の納付と共に出願人から異議申立てがなかった。